

# Interagency Regulatory Liaison Group Workshop on Reproductive Toxicity Risk Assessment\*

## Preface

On September 21–23, 1981, a workshop was held in Rockville, MD, to discuss specific issues related to the evaluation of data for risk assessment in reproductive toxicology (including teratology). The workshop was sponsored by the Interagency Regulatory Liaison Group (IRLG) and was organized by the IRLG Reproductive Toxicity Risk Assessment Task Group.<sup>†</sup> The Task Group's original charge was to develop criteria to support the consistent interpretation and use of reproductive and teratology data in the assessment of human risk by federal regulatory agencies. Early in the deliberations of the Task Group, it became obvious that a number of issues important to risk assessment were not well addressed in the literature. From these deliberations and from comments in response to a notice of the Task Group's work plan in the Federal Register (1), the workshop was convened to address specific issues that could influence the overall basis for policy-setting in reproductive toxicity risk assessment.

\*Editors: Carole A. Kimmel and Gary L. Kimmel, National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR. Present address: Reproductive Effects Assessment Group, U.S. Environmental Protection Agency, Washington, DC; and Vasilios Frankos, Office of Health Affairs, Food and Drug Administration, Rockville, MD. Present address: Environ Inc., Washington, DC.

<sup>†</sup>IRLG Reproductive Toxicity Risk Assessment Task Group: Vasilios Frankos, Chairperson (Food and Drug Administration, Rockville, MD; presently at Environ Inc., Washington, DC); Allen Heim, IRLG Surrogate Liaison (Food and Drug Administration, Rockville, MD); Charles Anello (Food and Drug Administration, Rockville, MD); Ann B. Brown (U.S. Department of Agriculture, Washington, DC); Kathleen M. Burke (U.S. Environmental Protection Agency, Washington, DC); Jerry Chandler (National Institute of Occupational Safety and Health, Rockville, MD; presently at Food and Drug Administration, Rockville, MD); Thomas F. X. Collins (Food and Drug Administration, Washington, DC); William Farland (U.S. Environmental Protection Agency, Washington, DC); Roseanne A. Hurwitz (Occupational Safety and Health Administration, Washington, DC); Peter Infante (Occupational Safety and Health Administration, Washington, DC); Carole A. Kimmel (Food and Drug Administration, Jefferson, AR; presently at U.S. Environmental Protection Agency, Washington, DC); Gary L. Kimmel (Food and Drug Administration, Jefferson, AR; presently at U.S. Environmental Protection Agency, Washington, DC); Franz Rosa (Food and Drug Administration, Rockville, MD); Theodora Tsongas (Occupational Safety and Health Administration, Washington, DC); Ann Wilk (Food and Drug Administration, Rockville, MD); Marilyn Wind (Consumer Product Safety Commission, Bethesda, MD).

The workshop was organized into six workgroups, each of which was given a specific area to address. Two workgroups, one in teratology and one in male and female reproductive toxicology, considered the animal and human endpoints which are useful for human reproductive risk assessment; e.g., the comparability of endpoints across species, the relationship of various endpoints to one another, the significance of transient effects, the relationship between maternal and fetal toxicity, the sensitivity of endpoints, and the endpoints which can be monitored in human epidemiology studies. Two other workgroups, again one in teratology and one in male and female reproductive toxicology, discussed the information available on mechanisms of action and its use in the interpretation of experimental data and risk assessment; e.g., the use of mechanistic data to help explain interspecies variation in response and to allow for development of more appropriate models for extrapolation of animal data to humans, the information on gene/toxicant interaction that might be useful for predicting risk, and the early indicators of toxicant effect that might be used to predict adverse outcomes. A fifth workgroup addressed the evidence that can be gained from pharmacokinetic studies in estimating potential reproductive and developmental risk; e.g., the process by which pharmacokinetic data can assist in choosing the appropriate species for testing and estimating risk, in selecting appropriate dosing regimens, in defining the target for exposure, in predicting thresholds, and in relating exposure to critical or sensitive times in development or in the reproductive cycles. The sixth workgroup was asked to address the procedures available for assessing risk from available human and animal data including: qualitative evaluation of study design and data, e.g., internal consistency of data, evidence of a dose–response relationship, and reproducibility of effects in multiple species; and quantitative evaluation of data, e.g., statistical procedures and mathematical modeling for low dose extrapolation versus the use of safety factors. The participants in each of the workgroups were selected with the aim of achieving a balanced discussion among representatives from academia, industry, and government.

The workshop provided a valuable forum for the expression of diverse scientific opinions and each workgroup provided a thoughtful evaluation of the difficult

issues they were asked to address. The panel chairpersons were asked to draft a document summarizing their deliberations which was then reviewed by members of the workgroup following the meeting. With the disbanding of the IRLG, there were a number of delays in finalizing the reports, but since all contained valuable insights and information which remain state-of-the-art, the Commissioner of the Food and Drug Administration decided to pursue the publication of the reports under the auspices of the FDA. The reports here represent updated versions of the original reports. They are not meant to be all-encompassing or to be a thorough review of the literature, but rather to state the consensus of opinion for each workgroup on the specific issues they were asked to address concerning use and interpretation of data for risk assessment.

Several other workshops and symposia have appeared since the IRLG workshop and may be consulted for more detailed information on methods and testing procedures (2-4). Recently, the FDA has completed a report detailing the requirements and recommendations each FDA Center has for studies to produce data in reproductive and developmental toxicology. In addition, the U.S. Environmental Protection Agency (EPA) has published proposed guidelines for risk assessment for developmental toxicity (4) and plans the development of guidelines for male and female reproductive toxicity in 1986. Another interagency group, the Interagency Risk Management Council (IRMC), is currently working to develop guidelines that can be used by all regulatory agencies for risk assessment in reproductive and developmental toxicity. The reports of the IRLG workgroups were very useful in the development of the EPA guidelines, and we hope they will continue to assist regulatory agencies in developing risk assessment policies as well as point to areas lacking in adequate knowledge so that research efforts can be developed to address these needs.

We would like to acknowledge the sponsorship of the Interagency Regulatory Liaison Group and the Food and Drug Administration which was the lead agency for this effort. We are indebted to the chairpersons of the workgroups for their success in focusing the discussions and developing the reports. All of the participants are

to be commended for their enthusiasm throughout the sessions which added immeasurably to the success of the workshop. We would like to recognize especially Ms. Janet Cunningham, Ms. Tina Sykes, Ms. Cynthia Hartwick, Ms. Rose Huber, and Ms. Tonya Richardson for their excellent clerical efforts without which the reports could not have been completed. Also, we would like to thank Dr. John F. Young and Dr. William Slikker Jr. for their assistance in editing and completing this document.

## **Workgroup on Endpoints of Teratogenicity (Developmental Toxicity)\***

### **Introduction**

In addressing questions posed under the heading "teratology," it is essential to identify direct effects of a test agent on the conceptus from indirect effects on the pregnant female. That is, to distinguish clearly the adverse effects on the products of conception from effects on specific adult target organs either directly or indirectly necessary for reproduction. This latter aspect is covered by the workgroup on Reproductive Endpoints and will not be discussed further here. Furthermore, before one can clearly discuss adverse effects on the conceptus, it is important to place the necessary terminology into an understandable context consistent with contemporary understandings. Both semantic and regulatory confusion exists regarding the terms, "teratogenicity" and "teratogen." Their meanings have shifted in some quarters in recent years, and the range of meanings from the strictest to broadest interpretations is now confusing to the point of use being counterproductive to understanding. In view of this, it was concluded that endpoints of toxicity applicable to the conceptus would best be referred to under the general term "developmental toxicity," which includes, as one of its parts, teratogenicity by its strictest definition; i.e., the production of grossly abnormal offspring in a specific experiment. However, any consistent dose-related adverse effect on any aspect of development would be worthy of consideration provided it occurred above the threshold of effect and background incidence of such effects in comparable control animals. The principal manifestations of disrupted developmental biology are: death of the conceptus, gross structural malformations, functional impairment, and altered growth and/or developmental patterns. These are not necessarily listed in a hierarchical order but, as far as can be determined, administration of a material at dosages capable of increasing the incidence of frank malformations usually will provoke other adverse effects as well. These effects can include any or all components of a spectrum of effects, i.e., fetal death, alteration in general growth pattern, increased incidences of minor alterations and developmental variations which may either be permanent or

\*IRLG Workgroup on Endpoints of Teratogenicity: E. Marshall Johnson, Chairperson (Jefferson Medical College, Philadelphia, PA); Carole A. Kimmel, IRLG Contact (National Center for Toxicological Research, Jefferson, AR); Allan R. Beaudoin (University of Michigan Medical School, Ann Arbor, MI); William R. Campbell (Ciba-Geigy Corporation, Greensboro, NC); Jane E. Harris (Food and Drug Administration, Washington, DC); Andrew G. Hendrickx (University of California, Davis, CA); Robert Hill (Syntex Research Inc., Palo Alto, CA); Benjamin Jackson (Food and Drug Administration, Washington, DC); Norman W. Klein (University of Connecticut, Storrs, CT); Granville A. Nolen (Proctor and Gamble Company, Cincinnati, OH); Anthony K. Palmer (Huntingdon Research Centre, Huntingdon, England); Franz Rosa (Food and Drug Administration, Rockville, MD); James L. Schardein (International Research and Development Corporation, Mattawan, MI); Robert E. Staples (E. I. duPont de Nemours and Company, Newark, DE).

transient in nature.

Adverse effects on a developing system do not necessarily occur as a strict continuum of responses in the sense that one type leads to the next or that one is invariably produced at a lower dose than another, but types of response can sometimes be viewed as a spectrum of effects. There are examples where breaks in the continuity of the spectrum have been observed, e.g., increased incidence of fetal death without an increase in the number of malformations (6). But, the possibility that closer examination would reveal such continuity, e.g., malformation preceding death (7), cannot be excluded in routine safety evaluations. Experimental studies can be designed to resolve questions such as this, but they cannot be considered essential for the detection of developmental hazards or estimation of risk because both *in utero* death and malformation can be valid indicators of developmental toxicity. A major factor in evaluating the relevancy of any developmental toxicity endpoint assay is the proximity of the dose causing effects in the conceptus to that dose causing maternal toxicity. This issue is addressed in great detail later in this report.

While it may be assumed that most test agents will not increase the incidence of malformations without also provoking other changes, the reverse argument may not apply. Indeed, growth retardation, as indicated by retarded weight and/or ossification, or increased incidences of minor morphological variations often occur without a corresponding increase in malformations. This is more common among materials primarily toxic to the dam than it is for agents with the conceptus as their primary target, i.e., substances which cause developmental toxicity at a small fraction of the dose toxic to an adult. Similarly, observation of a reduced degree of skeletal ossification (evidenced by alizarin staining) is usually interpreted as being related to general growth retardation, but each instance requires careful examination and cautious interpretation.

## Comparison of Endpoints in Humans and Animals

The developmental toxicity endpoints encountered in experimental animals do not and should not be expected necessarily to mimic those observed in humans exposed to the same toxicant. Similarly, the specific agent-related endpoints in humans are not always reproduced in experimental animals (8). However, adverse developmental effects have been detected in one or more species of laboratory animals as a result of exposure to essentially all chemicals or physical factors known to be developmental toxicants in humans. The absence of absolute uniformity of response is not surprising when one considers the many critical differences which exist between the conditions of human exposure and those for animal models. For example, differences in dosage, placenta, metabolism, pharmacokinetics, critical periods of development, duration of gestation, etc. (9) can

be expected to affect expression of developmental toxicity. Nevertheless, the production of adverse developmental defects in animal models lends support to the current view that such findings in experimental animals identify most chemicals that are potentially hazardous to human development (10,11). Since so many test agents manifest developmental toxic signs at, or very near to, the maternal maximum tolerated dose (MTD), one expects to encounter more effects in experimental animals than in man, where exposures tend to be less and where epidemiologic endpoint assays of toxicity are more difficult to ascertain and specifically relate to an agent.

## Levels of Concern Regarding Transient Effects

In routine tests for developmental toxicity, alterations of questionable biologic importance such as developmental delay are sometimes observed. In considering the significance of these or any other possible endpoints of developmental toxicity, the experimental data should be carefully examined for evidence of maternal toxicity. For example, a transient delay in fetal ossification or patterns thereof, produced only in fetuses of dams who are themselves manifestly affected by the treatment, are of questionable significance. In marked contrast, a permanent alteration in fetal development at some small fraction of the exposure needed to produce overt maternal toxicity would be of marked significance. To evaluate accurately the significance of some findings in relation to risk assessment, it may be necessary in some instances to demonstrate that a developmental delay is truly transient, and that the repair and innate self-regulation processes of the conceptus have not been exceeded sufficiently to interfere with normal function. An important area of future basic research would be determination of whether the offspring are more vulnerable to a second insult (by the same or different agents) during the period of a transient effect.

## Relationship between Adult and Developmental Toxicity

A major goal of testing for developmental toxicity is to determine whether a test substance is a greater hazard to the conceptus than it is to the pregnant female or adult male. As a general rule, an agent that causes detrimental effects in the conceptus at a dose level that also adversely affects the pregnant animal is considered to be of less concern and is a lower priority for detailed safety evaluation than an agent that detrimentally affects the conceptus at a dose level that is not harmful to the pregnant animal (12,13). There are notable exceptions, however, and in the human, special consideration should be given to agents considered acceptable by the mother (14,15) (e.g., smoking, alcohol consumption, life-saving drugs, employment in a high-exposure

environment). Even though adverse effects are produced in both the adult and the conceptus at the same general dosage, the adult may recover from the toxic exposure, but the embryo may be irreparably altered.

In evaluating the degree of risk to the human conceptus, one must carefully consider a number of differences that exist among animal species and between animals and man. These include genetics, metabolism, anatomy, physiology and other so-called "inherent" interspecies differences in development, or the anatomy and physiology of the placenta between experimental animals and man (16). These inherent factors can influence the outcomes of developmental toxicity testing. They need to be identified and their influence on test results should be clarified both for general influence and for modifying effects on the results produced by individual substances.

### **Interpretation of an Increased Incidence of Spontaneously Occurring Defects and Minor Variations**

It is recognized that in all animal species there is a detectable incidence of spontaneously occurring defects (17). Exposure- or dose-related increases in such defects in test animals are considered as manifestations of developmental toxicity, and are evaluated as being due to the test agent. Such increases are, for purposes of safety evaluation, as relevant as are dose-related increases in any of the other four classes of developmental toxicity endpoints.

Increased incidences of developmental variations (e.g., skeletal), with or without associated frank gross anatomical malformations, also may be present; they are interpreted to be indicators of developmental toxicity (18,19) when elicited in a dose-related manner at incidences significantly above comparable controls. If they are produced by exposures markedly below those inducing adult toxicity, the test substance should be considered as a potential hazard to the conceptus. However, some variation may represent temporary retardation of growth, development or degree of ossification, and the effect may be readily reversible with continued maturation (20). Such findings may merit less concern than would those of a more lasting nature.

### **Postnatal Assay of Developmental Toxicity Endpoints**

Tests for developmental toxicity should include postnatal endpoints that may be altered prenatally or during early postnatal development. The assessment of these endpoints can be incorporated into reproduction and/or developmental toxicity studies (21). The endpoints selected for evaluation will vary depending upon the nature of the test agent being tested, its use, and the amount of expected or actual human exposure. Reliable endpoints include survival, growth rate to maturity,

timing of selected developmental landmarks, feed consumption, efficiency of food utilization, and reproductive capability (22). Histomorphologic, hematologic, and clinical chemistry data may also be useful in some instances (23). As they become validated, additional endpoints of developmental toxicity may include measurements of neurobehavioral status, immunologic, respiratory or gastrointestinal function, or developmental enzyme patterns (21). Studies of the postnatal animal throughout its lifespan should be reserved for special products and/or problems, since they seldom yield information other than data relevant to evaluations of chronic toxicity, lifespan, or incidence of carcinogenicity.

### **Data Evaluation and Interpretation**

It is appropriate to group responses into some overall indicator of developmental toxicity, or to consider the incidence of "normal" offspring. Grouping of data from various types of endpoints has been done by a few investigators (24), and the interpretation of data may differ depending on the way in which data are grouped (25). The most appropriate method for evaluating data from experimental studies is yet to be determined, and this problem should be addressed by further study to aid the regulator in making appropriate assessments of risk for human development.

### **Epidemiologic Endpoints of Developmental Toxicity**

Developmental toxicity endpoints have been monitored inadequately by epidemiologic studies. To date, structured or formalized epidemiologic monitoring has not been the initial source of information revealing a developmental toxicity endpoint due to a specific agent, although recent expansion of epidemiologic studies of birth defects is increasing the probability that they will be able to do so in the future. Though actual cause and effect relationships sometimes may be difficult to establish, developmental toxicity endpoints in humans related to specific agents have been identified primarily by case reports. However, because the value of case reports in identifying human developmental toxicants has not been fully appreciated, case reporting has not been fully utilized. Identification of an endpoint depends on the uniqueness of the exposure and the uniqueness of the event, and association of exposures with events which are not unique or unusual is much more difficult. Observation of only a few unique events may be sufficient to establish an association (26), but the study of cohorts of exposure can only identify marked developmental hazards such as thalidomide.

In summary, new levels of understanding of the meanings and possible utility of developmental toxicity safety evaluations are emerging. They are predicated on classic principles of developmental biology and have evolved from concepts formulated by the pioneers in

experimental teratogenesis. Precision of terminology regarding the endpoint manifestations of altered development and a degree of consensus regarding its application constitute a major step in data interpretation. A pragmatic perspective on the concept that most anything can injure development if the dose or exposure level is high enough is achieved when developmental toxicity is related to adult toxicity. The determination of whether the conceptus is the "target" of a specific test agent or is only at risk secondary to, or concomitant with, adverse effects on the mother is a useful means for initial identification of developmental hazards (27). This allows a means for more accurately assessing risk. In addition, cross-species extrapolation can be made with even greater confidence when interspecies differences and/or similarities of pharmacokinetics are examined both for *in utero* and for postnatal developmental effects.

## Workgroup on Mechanisms of Teratogenicity\*

### Introduction

The mechanism of action of a teratogen, in the strictest sense, is the fundamental physical or chemical process which initiates a sequence of perturbed developmental events leading to an observable toxic response. In a broader sense, mechanisms of action have been defined at various levels: biochemical (molecular or sub-cellular), cellular, tissue, the embryo-placental unit, and the pregnant dam. In the determination of the components of an observed response, discrete or multiple mechanisms of action should be considered along with various levels of repair or regulation. Teratogenesis is a complex, poorly understood process, potentially involving perturbations of maternal-fetal, tissue-tissue, cell-cell, nuclear-cytoplasmic, and molecular interactions. Consequently, the mechanisms by which most known teratogenic agents act during embryonic or fetal development are only beginning to be defined. Hence, it may be tenuous to use such incomplete information as the basis for risk assessment of suspected hazardous agents given the complex nature of the information available at present.

Nevertheless, there are distinct advantages to be accrued from studies on mechanisms and these will even-

tually help in the evaluation, prediction, and assessment of risks. Through studies on mechanisms, a number of concepts have been clarified. These are mentioned briefly below, along with some examples, wherever appropriate.

## Critical Periods in Development and Resultant Syndromes

Environmental agents with known pharmacological actions have been used extensively as probes into embryonic or fetal development in laboratory animals. Chemicals such as alkylating agents or inhibitors of DNA or protein synthesis, that have cytotoxic activity have been employed to determine the critical time of embryonic organ susceptibility as well as the progressive susceptibility of various differentiating cells within a single discrete organ system. Limb development, for example, was found to be altered by exposure of mouse embryos to cytosine arabinoside, a potent cytotoxic drug, in a manner that reflected sensitivity along the proximal-distal axis of the limb. Thus, either upper, middle or distal (digits) segments were missing upon exposure to the teratogen at progressively older stages of development (28). Chemicals with either a narrow or a well-documented single biochemical site of action have been valuable in defining the role of specific endogenous substances in organogenesis.  $\beta$ -Aminopropionitrile, a lathyrogen that specifically inhibits collagen crosslinking, induces a high percentage of cleft palate in rat embryos when administered during a short period of time prior to palatal shelf elevation. This finding indicates the importance of this protein and the time when synthesis is critical for normal palatal development (29). Finally, both broad and specific agents have been used to investigate interactions which occur between developing organs and tissues. Because abnormalities induced by different classes of compounds can result in discrete organ deficits or syndromes, i.e., the fetal alcohol syndrome, fetal hydantoin syndrome, diamox (acetazolamide) syndrome in animals, an understanding of the relationships of normal inter- and intra-organ developmental sequences as well as the ability of reference teratogenic compounds to interfere with these processes can lead to insights regarding species differences in teratogenic response as well as the teratogenic potential of new chemicals.

## Dose Response

The study of the effects of increasing dosages of teratogenic agents has led to an improved understanding of how increased dose influences possible embryotoxic outcomes and how these outcomes putatively relate to one another. The four manifestations of embryotoxic responses are functional deficits, growth retardations, malformations and death of the embryo/fetus (30). One way in which the four manifestations of embryotoxicity may be related to each other is diagrammed in Figure

\*Devendra Kochhar, Chairperson (Thomas Jefferson University, Philadelphia, PA); Marilyn Wind, IRLG Contact (Consumer Product Safety Commission, Bethesda, MD); John DeSesso (Mitre Corporation, McLean, VA); Elaine Francis (US Environmental Protection Agency, Washington, DC); Casimer Grabowski (University of Miami, Coral Gables, FL); Gary L. Kimmel (National Center for Toxicological Research, Jefferson, AR); Guillermo Millicovsky (University of North Carolina, Chapel Hill, NC); William Scott, Jr. (Childrens Hospital Research Foundation, Cincinnati, OH); Richard Skalko (East Tennessee State University, Johnson City, TN); John Strange (Franklin Research Center, Silver Spring, MD); Ann Wilke (Food and Drug Administration, Rockville, MD); Ernest Zimmerman (Childrens Hospital Research Foundation, Cincinnati, OH).

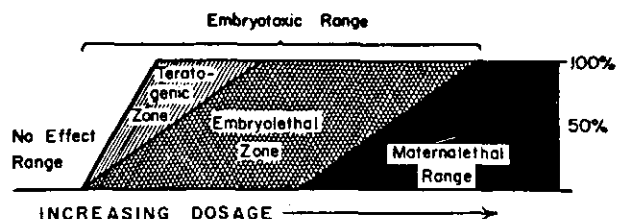


FIGURE 1. Diagram of the toxic manifestations shown by the embryos and the maternal organisms as dosage of a teratogenic agent increases. From Wilson (30) used with permission of Academic Press.

1. In this scheme, a low dosage of a given teratogen may elicit no observable embryotoxic response; with increasing dosage a teratogenic or other embryotoxic response may be observed and be considered the "threshold" dose for that response under the experimental conditions used. Further increases in dose may result in more severe embryotoxicity and/or maternal toxicity.

An example of an agent which has been studied for demonstration of a dose-response effect is hydroxyurea. Scott et al. (31) administered hydroxyurea to pregnant rats on gestational day 12. The lowest dose (250 mg/kg) used produced no malformations. Intermediate doses (500 or 750 mg/kg) produced increasing amounts of fetal malformations; the highest dose (1000 mg/kg) resulted in a higher incidence of embryoethality. In addition, Butcher and colleagues (32) have demonstrated behavioral deficits in rat pups whose mothers were treated with low levels of hydroxyurea (375 mg/kg) at the same time in gestation (day 12). A similar spectrum of embryonic responses at increasing doses has been obtained with other agents, including cytosine arabinoside (28,33).

## Species/Strain Differences

Species and strain differences in teratogenic response are the rule rather than the exception. Studies of mechanisms of teratogenic action in the broadest sense have helped us to understand that basic species differences can occur in the mother, the placenta, or in the embryo itself.

**Mother.** In the mother, veratrum alkaloids produce a syndrome of craniofacial defects when administered to sheep during early gestation. When they were administered to rats or rabbits during an equivalent gestational stage, these agents were totally ineffective. However, if gastric contents were made alkaline, these alkaloids induced a similar array of structural malformations in rabbits (34).

Chernoff (35) has shown that different strains of mice respond to ethanol with different degrees of developmental toxicity. He has demonstrated a close positive correlation between maternal blood level of ethanol and developmental toxicity between strains and further solidified this idea by demonstration of a negative corre-

lation between the alcohol dehydrogenase activity and developmental toxicity.

Wilson et al. (36) demonstrated that aspirin is more teratogenic in rats than in monkeys. Following administration of equivalent dosages based on maternal weight, total maternal plasma levels of the major metabolite, salicylic acid, were the same in both species. However, the amount of salicylic acid in the embryo was much greater in the rat and, in either species, levels in the embryo compartment of salicylic acid were nearly identical to the level of unbound salicylic acid in maternal plasma. Thus, plasma binding in the mother was held responsible for the difference in species response to aspirin.

**Placenta.** Trypan blue is teratogenic in rodents which possess and utilize a yolk sac placenta during early organogenesis. The yolk sac provides histiotrophic nutrition during early rodent development, and trypan blue interferes with this process and is thought in this way to induce abnormal development (37). Species without a yolk sac placenta are presumably insensitive to trypan blue, perhaps due to a lack of inhibition of nutrition from sources other than the yolk sac.

**Embryo.** Different strains of mice respond with different frequencies of cleft palate to glucocorticoids. The level of glucocorticoid receptors in palatal cells correlates with cleft palate frequency in most mouse strains (38,39).

Carbonic anhydrase inhibitors in most strains of rats and mice produce a unique limb malformation syndrome characterized by a postaxial reduction deformity of the digits (40). For many reasons, the inhibition of carbonic anhydrase is thought to be responsible for this unusual malformation syndrome (41). Monkeys are insensitive to the teratogenic effects of these agents. Measurement of carbonic anhydrase activity during the gestational stage of presumed sensitivity indicated that embryos of this species have little if any carbonic anhydrase and this factor could be the basis of species insensitivity to those drugs which are thought to act by inhibiting this enzyme (42).

## Structure/Activity Relationship

A number of agents that have structural similarity to thalidomide have been tested for teratogenic activity in appropriate species. Thus, we know many of the structural requirements for thalidomide teratogenesis. This knowledge has been useful in identifying potentially dangerous drugs and chemicals in our environment such as glutethimide (Doriden), Captan, and Folpet. We now know that the agents that do not possess the requisite structural specificity are not able to induce thalidomide-like teratogenicity. Conversely, new drugs have been synthesized which retain the favorable sedative properties but have abolished the teratogenic properties of thalidomide, a concept which could be helpful with other teratogens.

## Drug Interactions

When a combination of agents is applied to a pregnant animal, the results can be either protective, synergistic, or have no additional effect. An example of synergism reported in the study of Ritter et al. (43) is that treatment of pregnant dams with both an inhibitor of DNA synthesis and an inhibitor of RNA synthesis produced a greater incidence of developmental toxicity than either regimen alone. In contrast, treatment of pregnant dams with two different DNA synthesis inhibitors did not produce any additive or synergistic effect. Other examples of synergism include treatment of pregnant rabbits with compound 4880 and hydroxyurea (44) and treatment of pregnant rats with caffeine and acetazolamide (45).

Examples of protection include the cotreatment of pregnant rats with exogenous pyrimidines (especially deoxycytidine) and hydroxyurea or cytosine arabinoside resulting in virtually complete protection (28,46). Others have used folic acid "rescue" to protect or activate the embryotoxic effect of methotrexate and other folate antagonists (47). The antioxidant propyl gallate has been reported to ameliorate the teratogenic effects of hydroxyurea (48).

## Summary

Since no well-defined mechanism of teratogenic action of any agent is known at present, it is difficult to predict the risk entailed from exposure to a suspected developmentally toxic agent. However, chemicals which interfere with nucleic acid integrity have been studied well enough to allow one to make reasonable predictability of risk. For other classes of compounds, such as CNS depressants and vasoactive agents, one could reasonably predict their teratogenic potential on the basis of empirical (non-mechanistic) observations. Reliable risk assessment requires further studies on the mechanisms of action of agents at all levels (biochemical, cellular, tissue, embryo-placental unit and pregnant dam) before this factor can be used as a basis for realistic predictability.

The study of mechanisms is important in our understanding of the degree of concordance between test an-

imals and man and, therefore, should be able to improve the extrapolation of test findings to potential toxicity to the human conceptus. The concepts that have been reviewed are characteristic of issues within teratology that have been clarified through mechanistic studies.

Understanding developmental sequences, their attendant timings, and the mechanism by which teratogens perturb the process, should permit both critical and safe periods to be identified which would allow more effective assessment of teratogenic risk.

Teratogenic responses to a compound may vary according to the route of exposure. An understanding of the mechanisms involved in activation, inactivation, and pharmacokinetics may allow the optimal delivery of a compound to populations which require treatment in pregnancy and thus minimize teratogenic risk.

The understanding of structure-activity relationships coupled with a better understanding of the compound's mode of action in the maternal and embryo/fetal organisms should allow for the development of pharmaceutical and other compounds which would pose a lower risk to human females during pregnancy.

Model systems should be devised to determine the causes of teratogen-induced malformations. Human populations which are at high risk for general teratogens and for specific teratogens to which they are exposed may be identified. Current research in recombinant DNA technology offers the promise of altering gene structure in human offspring. If underlying adverse teratogenic responses in a population can be identified, then it may be feasible to alter gene structure to prevent the consequence of inherent error.

## Workgroup on Endpoints of Reproductive Toxicity\*

### Introduction

Reproductive toxicity deals with the effects of toxicants on adult reproductive function and development of the offspring which may be produced by alteration of a wide range of processes in either the female or male. These processes include those associated with the primary and accessory sexual organs and with fertilization, as well as those which impact more indirectly on normal reproductive function; e.g., neuroendocrine control, general physiological and psychological health, and nutrition. Following fertilization, processes associated specifically with pregnancy are also vulnerable; e.g., implantation, placental formation and function, conceptual development, and parturition and lactation. The focus of this workgroup was on the entire reproductive process, although coverage of the developing embryo/fetus was limited, since this was the purview of another workgroup. To detail each of the vulnerable processes and how they may be altered is beyond the scope of this discussion. Several references are available which have discussed these in greater detail (2-4). Specific exam-

\*Vilma Hunt, Chairperson (Pennsylvania State University, University Park, PA); Kathleen M. Burke, IRLG Contact (U.S. Environmental Protection Agency, Washington, DC); Rajender Abraham (Albany Medical College, Albany, NY); David Anderson (U.S. Environmental Protection Agency, Washington, DC); Mildred S. Christian (Argus Research Laboratories, Inc., Perkasie, PA); T. F. X. Collins (Food and Drug Administration, Washington, DC); Robert Diener (CIBA-GEIGY Corporation, Summit, NJ); Brian D. Hardin (National Institute of Occupational Safety and Health, Cincinnati, OH); Richard M. Hoar (Hoffman LaRoche, Nutley, NJ); Peter F. Infante (Occupational Safety and Health Administration, Washington, DC); Charles Keller (Exxon Corporation, East Millstone, NJ); Richard J. Levine (Chemical Industry Institute of Toxicology, Research Triangle Park, NC); James H. Spraul (Monsanto Company, St. Louis, MO).



ples have been included where they provide clarity both in this report and in that on mechanisms of reproductive toxicity.

Within each of the processes of normal reproductive function, the various events which may be altered and lead to a toxic response can represent a continuum, in the sense that various endpoints are integrally related to other biological processes that may be the specific target of toxicity. For example, an agent may alter male reproductive function by affecting spermatogenesis directly in the testis or by affecting hormone production at the level of the pituitary/hypothalamus, or indirectly, by altering normal function of the sex accessory glands. In any case, the ultimate result is the lack of fertile sperm interacting with the oocyte, and consequently, reproductive failure. Therefore, it is important that these continuums are recognized in establishing and assessing reproductive endpoints in order to appreciate the complexity of a particular response and better understand their applicability in assessing risk.

## Specific Endpoints

As there are a wide range of reproductive processes that may be altered following insult, there are a comparable number of endpoints which potentially can be used to evaluate toxicity. Ultimately, the most sensitive endpoints may reside in or be closely associated with the target of insult. Currently, however, risk assessment must rely on data from laboratory testing or from epidemiological studies, which for the most part assess only the final outcome of the reproductive process. The following outlines some of the endpoints currently in use or under development for reproductive toxicity screening

***In the Male.*** Semen characteristics are parameters including sperm count, motility and morphology, and seminal volume. They are useful endpoints for characterization if techniques are standardized relative to abstinence time, collection and counting procedures, etc. Nevertheless, certain limitations of the parameters must be recognized, both biological and sociological. For example, the number of motile sperm is a sensitive indicator of human fertility. However, this may be a less sensitive indicator in laboratory animals, due to their tendency to produce sperm in considerable excess over that required for normal reproductive function. In addition, motility and morphology are generally subjective evaluations, and controlled techniques for uniform scoring are only beginning to be established. Human subjects may be reluctant to participate or be unwilling or unable to satisfy the requirements of the study design.

*In vitro* oocyte penetration using zona-free hamster ova and human sperm (49), appears to be a highly sensitive and specific assay for male fertility in the population tested. It is the only functional measure available for sperm-egg interactions and may reflect subtle changes in the standard semen parameters by indicating reduced penetration of eggs. However, it is perhaps

most meaningful in indicating subtle changes in sperm function (ability to fuse with and penetrate the ovum) in the absence of any change in the standard semen parameters. Additional validation of the technique is required. Technical expertise required for the test currently prohibits its widespread use.

Preliminary data indicate that there may be marker enzymes in the acrosome with applicability to fertility testing. These enzymes may be necessary for sperm penetration into ova. However, these tests are in the early developmental stage.

New tests using flow cytometry and fluorescent staining (F-body staining) of chromosomal changes (chromatin stability and nondisjunction) may provide information on genetic or chromosomal damage in sperm. These methods have not yet been validated as predictors of fertility or production of viable offspring.

Endocrine profile tests determine the blood levels of pituitary hormones which exert a control over testicular function, and of testosterone which conversely exerts a negative feedback on the pituitary. Thus, these tests may indicate the level of alteration in the pituitary-testis axis. Presently, however, they demonstrate changes in fertility only at extreme levels. Additional comparison and validation are required before they can significantly aid a testing program.

Cervical mucous penetration is a functional test primarily measuring sperm motility and the ability of sperm to migrate through the cervix. The test is not complicated and can be done quickly. It does not indicate the fertilizing capacity of sperm, but focuses on sperm transport. The test is only slightly better than standard semen analysis in evaluating fertility.

*In vivo* fertility testing in laboratory animals includes endpoints which are analogous to those occurring in the human, and they can provide an important indication of reproductive toxicity. Nevertheless, species differences (e.g., excess sperm production in animals, noted above) limit data extrapolation to the human unless the test design and data analysis adjust for these differences. The use of fertility as an endpoint in human studies has been more retrospective and is limited by the low conception rate in the human. Moreover, the couple is a biological unit and therefore a defect or abnormality in the female may mask or hinder detection of a problem in the male. Thus, it often is difficult to identify the infertile partner unless either partner has a prior history of fertility.

Tests designed to evaluate the outcome of pregnancy focus on endpoints such as spontaneous abortion, birth weight, live/dead births, birth defects, and neonatal development and survival. Such tests provide an overall assessment of reproduction, including libido, fertility and pre- and post-natal development. In laboratory animals, the male contribution to a toxic response is not easily established, since the female contribution cannot be excluded entirely. It is also necessary to ensure that the entire male cycle is considered in the treatment period prior to cohabitation (usually 60 days; should be



80–90 days) and that a sufficient number of male animals be tested. Actually, the current study designs give emphasis to the female (11,50). In the human, data are often retrospective, and postnatal observations may not be apparent for a year or longer. These measures may be meaningful on a population basis but are not always apparent or valid in individual cases; separation of the male and female contribution to the reproductive dysfunction requires careful study design. A more detailed discussion of approaches in human studies can be found in the risk assessment report.

**In the Female.** Oocyte and follicular toxicity is a quantitative assay in which oocyte and follicular number are determined following agent exposure (51). The experimental animals, usually mice, are sacrificed at intervals after exposure, and the ovaries are prepared and examined microscopically. It appears that the assay is more sensitive to exposure than standard fertility endpoints, since a considerable reduction in oocyte/follicular number must occur before fertility is actually altered. However, this procedure is extremely time-consuming and requires an advanced level of expertise both in technique and evaluation.

As in the male, endocrine profiles include the blood levels of pituitary and gonadal hormones which exert control over reproduction. Presently, there are both *in vivo* and *in vitro* approaches for measuring these hormones individually. However, direct correlations between alterations in a specific hormone level and fertility following agent exposure have not been established. Other events important to overall reproduction (e.g., parturition and lactation) are also under hormonal control, and alterations in them could potentially be estimated by measuring blood levels of specific hormones. Nevertheless, the validity of this type of association has not been demonstrated.

Endpoints which are derived from standard *in vivo* reproduction studies continue to be most accepted in assessing reproductive toxicity (11,50). For example, the fertility index represents matings which result in pregnancies and indicates the ability of the female to become pregnant. Viability and growth indices can be indicators of agent-induced alterations in lactation, postnatal nourishment or maternal care of the offspring. Developmental landmarks such as vaginal opening and onset of the estrous cycle can also indicate alterations in normal hormonal balance. As indicated above, these tests can provide important information relative to male reproductive function.

## Significance of Transient Effects

A number of factors, including drugs, fever, or stress, are known to cause transient effects on several of the reproductive parameters. But, an important concern relative to risk assessment is the significance of such transient effects. Transient changes (i.e., toxic manifestations of relatively short duration in which full functional recovery occurs after cessation of exposure) by

definition may not have any major biological significance; however, long-term or additive effects of transient changes are possible. Transient effects observed in reproductive toxicity studies, e.g., temporary fluctuations in sperm characteristics, fertilizing capacity, or ovum production and transport, may represent one point on the dose–response curve; higher exposure levels may produce irreversible effects. Thus, such transient effects should be considered in the initial experimental design, data collection and analysis. In general, a defined approach to assessing the impact of transient events is not available. For the most part, we still must rely on good scientific judgment when estimating the potential health hazard of an agent whose primary toxic effects are transient.

## Comparison of Endpoints in Humans and Animals

Currently, the paucity of information for both experimental animals and humans does not allow the correlation of reproductive effects between species. Therefore, any expected similarity of toxic responses will be based primarily on general assumptions regarding the similarity of biological processes. Certainly, many of the reproductive processes described in the laboratory animal appear to have correlates in the human. The basic processes of gamete development and transport, fertilization, and implantation are similar, as is overall neuroendocrine control. The need for additional comparative data is obvious. When exposures of humans to potentially toxic substances have occurred, efforts should be made to evaluate the effects by epidemiologic studies (see below and report on risk assessment). However, because human exposure data are limited, we still must rely primarily on animal studies. This points up the importance of developing animal models for the human situation. In the absence of specific data to the contrary, adverse effects in experimental animals should be presumed to indicate a potential risk to human reproduction.

## Epidemiology

In order for epidemiological studies to be useful in reproductive risk assessment, it is necessary to obtain measurements of both exposure and health effects. Measurements of exposure should include dose (several doses if possible to establish a dose–response relationship), route of exposure, duration of exposure, time of exposure relative to conception, and health outcome of interest. In human epidemiological studies, such information may not be available, and estimates of dose may have to be determined from biological monitoring data, duration of employment, description of job duties, or work area.

Tools that have been useful in assessing reproductive endpoints in human epidemiological studies include use of the questionnaire to obtain information on reproduc-

tive outcome and potential confounding factors, physical examination of the reproductive system of either parent or examination of the product of conception, laboratory measurement (such as sperm count), and histopathological examination of the reproductive system of either parent or the offspring.

Human epidemiological studies are subject to several limitations. As noted previously, precise exposure information is more difficult to obtain for humans than in animal studies. In addition, confounding factors cannot always be excluded or controlled, making interpretation difficult at best. Assessment of reproductive endpoints is also affected by the necessity of voluntary participation, privacy considerations, religious objections, reliability of reporting and incompleteness of medical records. Histopathological data (e.g., testicular biopsies or fetuses in various stages of development) are also difficult to obtain. In spite of these limitations, epidemiological studies are important in establishing a relationship between exposure and reproductive effects in humans. A more detailed coverage of this area is presented in the report on risk assessment.

## Workgroup on Mechanisms of Reproductive Toxicity\*

### Introduction

The mechanism of action of a reproductive toxicant can be defined as the molecular interaction by which that chemical perturbs an underlying reproductive or developmental process. Although exact mechanisms have not been precisely established for a variety of morphologic, biochemical, or functional lesions resulting in reproductive dysfunction, possible sites of action can be visualized. Table 1 lists the variety of biological processes that are susceptible to toxic insult. This is compounded by the number of male and female target systems that may be affected. Normal reproductive function is dependent on the neuroendocrine system for the maintenance and function of the gonads and accessory sex organs. The gonads and accessory sex organs, in turn, carry out major roles as producers of germ cells, steroid hormones and an environment conducive to pro-

Table 1. Various reproductive functions susceptible to toxic chemicals.

Process	
Hormonal	Endocrine cell
	Hormone synthesis
	Hormone intracellular transport and release
	Feedback mechanisms
	Hormone transport and metabolism
	Transport
	Metabolism
	Hormone-mediated response
	Membrane transport
	Cytoplasmic receptors
	Nuclear translocation
	Chromatin interactions (DNA included)
	mRNA
	Protein synthesis
Cellular	Structural integrity of protein
	Functional integrity of protein
	Maintenance
	Excretion
	Metabolism
	Replication
	Differentiation
Secretory	Growth
	Ion transport
	Protein
Smooth muscle function	Microvasculature
Neurobehavioral	Tubular function

creation. It is beyond the scope of this discussion to detail each of these systems which have been reviewed well elsewhere (52-54). Rather, the discussion will focus on several concepts which have been addressed through mechanistic studies.

### Receptor Mechanisms

Many reproductive toxicants are likely to act in a fashion similar to endogenous reproductive hormones which initiate their action through a membrane or intracellular receptor. In such cases, pharmacokinetic parameters would determine the amount of agent which reaches a receptor to produce the toxicodynamic effect resulting in toxicity or some undesired physiologic response. Likewise, mechanistic studies at this level would provide: identification of the cellular target for the toxic agent at the molecular level; characterization of the ultimate form of the toxic agent studied; and elucidation of the agent-receptor interaction at the molecular level, including the short- and long-term consequences of that interaction.

The best example of toxicants which appear to have this common mechanism are the estrogenic xenobiotics, such as DDT and kepone (55,56). These have been shown to bind to estrogen receptors and stimulate estrogenic responses in several systems. The estrogenic activity is considered to be disruptive to reproductive capacity because it interferes with normal hormonal and developmental events. The mechanism of estrogen action is generally considered to involve the binding of

\*James Clark, Chairperson (Baylor College of Medicine, Houston, TX); William Farland, IRLG Contact (Environmental Protection Agency, Washington, DC); Robert L. Dixon (National Institute for Environmental Health Sciences, Research Triangle Park, NC); Jill D. Fabricant (University of Texas Medical Branch, Galveston, TX); Jerry L. Hall (University of North Carolina Medical School, Chapel Hill, NC); Charles A. Keller (Exxon Corporation, East Millstone, NJ); James C. Lamb (National Institute for Environmental Health Sciences, Research Triangle Park, NC); Thomas Marks (The Upjohn Company, Kalamazoo, MI); Donald Mattison (National Institute of Child Health and Human Development, Bethesda, MD); Martha Radtke (University of Cincinnati Medical Center, Cincinnati, OH); A. Louis Shor (Smith Kline Animal Health Products, West Chester, PA); James Stevens (Ciba-Geigy Corporation, Greensboro, NC); George M. Szczech (Burroughs Wellcome Co., Research Triangle Park, NC).

estrogen to cytoplasmic macromolecules called receptors. Estrogen receptor complexes undergo translocation to the nucleus where they bind to acceptor sites on chromatin. This nuclear binding is thought to stimulate transcriptional events which result in elevated RNA and protein synthesis. If estrogenic toxicants act in a similar fashion, then screening tests which could detect and classify these agents could be developed. For example, simple competitive inhibition assays for estrogen receptor binding can establish readily whether or not an agent binds to receptors. If binding occurs, then the agent may act as an estrogen agonist or antagonist, disrupting normal endocrine control.

Estrogen receptors have been described in a variety of tissues, including the hypothalamus, pituitary, and ovary, and a common mechanism of cellular action has been proposed for each (57). Therefore, an estrogenic insult during the perinatal or adult period may act at a number of biological targets. For example, an estrogenic compound could interact at the hypothalamic level to disrupt the mechanisms which control normal cyclic secretion of gonadotropins in adult life (58).

The knowledge that a compound interacts with a specific receptor makes it very likely that a receptor-mediated event is linked to the toxic action. This knowledge should ultimately provide new insights into reproductive toxicity testing and risk assessment. For example, many cellular and molecular processes tend to be similar in different species; thus, clear definition of the receptor for a toxic agent and its interaction with chemicals should aid interspecies comparisons of toxicity. Likewise, since many receptors influence the structural and regulatory integrity of genes, altered protein synthesis or abnormal regulation of protein synthesis may be early indicators of potential reproductive dysfunction that is initiated through a chemical-receptor interaction.

### Hypothalamic-Pituitary Mechanisms

A toxic chemical may adversely affect reproduction by altering the rate of secretion of one or more hormones that are synthesized and released by the hypothalamus or anterior pituitary gland (58). Of the hormones that are secreted by the anterior pituitary gland, the gonadotropins (luteinizing hormone, follicle-stimulating hormone, and prolactin) are most closely associated with reproduction. The gonadotropins control ovarian and testicular function, including steroid hormone secretion, follicular development and ovulation, and spermatogenesis. Hence, if gonadotropin secretion is suppressed, either by direct action on the pituitary or by suppression of hypothalamic-releasing factors, gonadal function is suppressed.

Alternatively, a toxicant could stimulate the secretion of prolactin, and the resulting hyperprolactinemia might suppress gonadotropin secretion. Prolactin secretion can be stimulated by substances that have estrogenic activity, substances that act as dopamine antagonists,

substances that inhibit dopamine secretion by hypothalamic dopaminergic neurons, and substances that cause hyperplasia of prolactin-secreting cells. Some of these actions of toxicants can be assessed by quantifying gonadotropin and prolactin secretion, whereas others such as releasing factors cannot be evaluated in a quantitative sense.

It is conceivable that toxicants with neurotransmitter and gonadotropin-like activity will be identified which mimic or antagonize the normal secretion of gonadotropins. Such observations form a basis for further work on mechanisms in this field and may in the future lead to insights concerning risk assessment.

### Inhibition of Steroidogenesis

Estrogens (primarily 17 $\beta$ -estradiol), progesterone, 17 $\alpha$ -OH progesterone, androstenedione, and testosterone are the predominant steroids produced by the human gonads during the reproductive years (52). They regulate gonadotrophin secretion and the reproductive cycle, as well as influence the development of the accessory reproductive organs and secondary sex characteristics. Regulatory steps in gonadal steroid secretion include substrate (cholesterol) availability, luteinizing-hormone induction of the 20, 22-hydroxylase-desmolase steps converting cholesterol to pregnenolone, and follicle-stimulating-hormone induction of granulosa cell aromatase activity converting thecal androgens to estrogens. If, at any point, the series of events leading to the synthesis and secretion of active steroids is disrupted, then control of the reproductive system is compromised. A toxicant may not demonstrate inhibition of steroidogenesis *in vitro* and yet be active *in vivo*, if it selectively affects gonadotropin-mediated events *in vivo* or progesterone synthesis stimulated by human chorionic gonadotropin. Hence these effects would be detectable only *in vivo* during a conceptive cycle. Similarly, agents affecting prostaglandins which induce luteal regression in some species may be active only *in vivo*. The agents may not act directly on the steroid-secreting cell, but indirectly by selective ovarian venoconstrictive action. Similar processes occur in the testis which lead to testosterone biosynthesis by the Leydig cells and may adversely affect germ cell maintenance by Sertoli cells, germ cell development, and accessory sex organ function. Despite these possibilities, most known inhibitors of steroidogenesis act by affecting specific enzymes in the steroid pathways (Table 2).

### Reproductive Toxicants

**Polycyclic Aromatic Hydrocarbons.** The polycyclic aromatic hydrocarbons (PAH) are ubiquitous environmental pollutants produced by combustion of fossil fuels. They are contained in automobile exhaust, smoke stack emissions, and cigarette smoke. Although these compounds have long been known to be toxic and car-

Table 2. Agents that inhibit steroidogenesis.

Steroidogenic step	Inhibitor
20 $\alpha$ -Hydroxylase	Amino glutethimide phosphate
Side chain cleavage	3-methoxybenzidine
Dehydrogenase, 3 $\beta$ -hydroxy- $\Delta$ 5-steroid	Cyanoketone
	Estrogens
	Azastene
	Danazol
Aromatase	4-acetoxy-androstene-3,17-dione
	4-hydroxy-androstene-3,17-dione
	1,4,6-androstatriene-3,17-dione
11 $\beta$ -Hydroxylase	Danazol
	Metyrapone
	SKF-12185
21-Hydroxylase	Danazol
17 $\alpha$ -Hydroxylase	Danazol
	SU-9055
	SU-8000
17,20-Lyase	Danazol

cinogenic, recent studies have also demonstrated reproductive toxicity.

Several PAHs have been demonstrated to destroy oocytes in weanling and sexually mature rats and mice (59). Treatment of a pregnant female will also destroy oocytes in the female fetus *in utero*. The mechanism of action is thought to depend on metabolism of the parent PAH to a chemically reactive intermediate which binds covalently to cellular macromolecules destroying the oocytes. The effect of oocyte destruction is to produce premature ovarian failure in the treated animals. Recent experiments have also suggested that certain PAHs can alter the ability of oocytes to complete meiosis, suggesting another mechanism for reproductive failure after ovulation.

Susceptibility of human or nonhuman primates to oocyte destruction by PAH is not clear. However, it has been demonstrated that smoking produces a dose-dependent decrease in the age of spontaneous menopause (60). Women smoking one or more packs of cigarettes per day have menopause approximately two years before nonsmokers. Women smoking half a pack of cigarettes per day have a median age of menopause about one year before nonsmokers. It has been suggested that the effect of smoking on the age of menopause is due to oocyte destruction by PAH from cigarette smoke. Many studies have suggested that cigarette smoke and nicotine can impair reproduction and fetal development in experimental animals as well as humans. It is not known if these adverse effects result from PAHs, nicotine, carbon monoxide, or protein pyrolyzates, all known constituents of cigarette smoke.

**Alkylating Agents.** Alkylating agents are useful in both the chemical industry and as therapeutics because of their chemical reactivity. In the chemical industry, they are used in a broad spectrum of synthetic reactions, and as therapeutics they are useful in treating neoplastic and some nonneoplastic diseases. The alkylating agents are also interesting because they represent known reproductive toxicants.

One of the first alkylating agents used therapeutically, busulfan, is known to produce gonadal failure in humans and experimental animals (61). Also, high-dose intermittent therapy for neoplastic disease with cyclophosphamide alone or in combination with other drugs destroys oocytes in an age- and dose-dependent manner. Other antitumor drugs also impair fertility and cause reproductive dysfunction in both the male and female (61).

Male antispermatogenic effects have been associated with various alkylating agents, for example, 1,2-dibromo-3-chloropropane (62-64) and its metabolites epichlorohydrin and  $\alpha$ -chlorohydrin (64), cyclophosphamide (65,66), and ethyl methanesulfonate (EMS) (67). These compounds act directly on the genetic material by alkylating the DNA (68). In the case of EMS, a specific nucleotide (guanine) is alkylated, and the genetic message is permanently misread. This leads to incorrect DNA and RNA replication and the subsequent synthesis of inappropriate protein sequences. These deficiencies result in cytotoxicity to the spermatogenic cells. Other electrophilic chemicals may also act directly on the DNA of spermatogenic cells and lead to cytotoxicity or improper genetic messages by mutagenesis or clastogenesis. The ability to predict interspecies responses to alkylating agents is dependent upon the general distribution and metabolism of these compounds in the various biological systems.

Of particular interest is the response of the reproductive system to cyclophosphamide (69). The mechanism of reproductive toxicity of cyclophosphamide results from metabolic activation of the parent compound and formation of reactive intermediates. Detoxification of reactive metabolites occurs through conjugation. Repair of resulting cellular damage after cyclophosphamide treatment requires replacement of damaged macromolecules and DNA repair.

An interesting difference in age-dependent sensitivity for gonadal toxicity from cyclophosphamide exists between males and females. Young females are more sensitive to oocyte destruction than older females. In males, however, the young (prepubescent) animal is more resistant to testicular toxicity than the older male. This difference in sensitivity represents differences in both reproductive and toxicological mechanisms. The rate of spermatogenesis in the prepubertal testis is very low, and since spermatogenesis is sensitive to cyclophosphamide, suppression of sperm cell division will provide some degree of protection against toxicity, consistent with the observed age-dependent differences in testicular toxicity. In the female, however, the most sensitive population appears to be the resting or small oocytes. As these are present throughout most of the life of the female in the same metabolic state, this cannot account for age-dependent differences in sensitivity. There is evidence, however, that the availability of pathways for detoxification of the reactive metabolite(s) of cyclophosphamide change with age, and that these changes parallel the observed changes in sensitivity.

Therefore, the opposite age-dependent changes in ovarian and testicular sensitivity to cyclophosphamide can be explained on the basis of an understanding of both reproductive biology and toxicology. In the testis, repopulation of seminiferous tubules will occur if spermatogonia remain after cessation of cyclophosphamide toxicity. In the ovary, however, since oogonia are not present after birth, oocyte destruction is permanent and irreplaceable. Cyclophosphamide has also been demonstrated to impair meiosis in rat oocytes.

**Solvents.** Recent studies have demonstrated that certain intermediate solvents, effective in creating mixtures of chemicals which are generally soluble only in either organic or aqueous solutions, may be reproductive toxicants. These compounds, such as the various glycol ethers, are very important industrial chemicals. Selected members of this class of compounds have adverse effects on testicular function in laboratory animals, while other members of the class are less active (70). The testicular toxicity of ethylene glycol monomethyl ether is the result of metabolism of the parent compound to the active toxicant, methoxyacetic acid (71). Other compounds in the class, such as ethylene glycol monobutyl ether and the propylene glycol alkyl ether series seem to be considerably less toxic to the reproductive system.

The structure-activity studies of such compounds are essential for safety assessment and for the elucidation of the mechanism of action. Unlike the alkylating agents, the glycol ethers seem to have little or no mutagenic activity. The mechanism of action seems to be an effect specifically on the late stage spermatocytes in the testis, leading to severe changes in testicular morphology, decreased sperm production, adverse effects on fertility, and very modest evidence of dominant lethality of these compounds (72). Compounds like the glycol ethers are apparently metabolized to an active metabolite, but the present evidence does not support a genetic or a hormonal mechanism of action. It is important to follow up the original toxicity and metabolism studies to identify the cellular target.

## Mechanism Studies and Risk Assessment

Although information on mechanisms would certainly be of value for interpretation of experimental data and risk assessment, from a practical point of view, a decision on human reproductive risk would most likely

have to be made long before such information is available. It often takes years of research effort to come up with a possible mechanism of action—not only toxicological but even pharmacological—and there are many compounds in widespread use whose mechanism(s) of action have not been determined. Drugs, pesticides, herbicides, etc., usually enter the marketplace with some information as to their physiological effects from which one can obtain a starting point for studying their mechanisms of action and potential toxicity. However, chemicals are often not developed with living organisms as a target, thus limiting the amount of data available. Frequently, studies at the molecular or biochemical level have not been carried out. There is no question that knowledge of the *in vivo* effects of chemicals on reproduction is valuable. However, it is highly unrealistic at this time to insist that the mechanisms of toxicity be ascertained before marketing the chemical.

We do not fully understand all of the molecular events of the normal reproductive pathways. Thus, it can be expected that the study of toxic mechanisms, many of which may affect several pathways, will be difficult. The development of new assays will be required as well as an upgrading of the training and expertise in this area.

Given the present state of the sciences of reproductive biology and toxicology, it is difficult to predict differential species susceptibility, even assuming knowledge of mechanism of action. However, in the long run, understanding mechanisms of action provides the greatest potential for prediction of reproductive hazards and risk assessment across species. Although mechanistic data may be helpful and should be considered as it becomes available, there are no general recommendations as to how such data should be used in hazard and risk analyses. Such recommendations may be possible in the future when an integration of pharmacokinetic and toxicodynamic parameters result in insights from which predictions can be made.

## Workshop on the Role of Pharmacokinetics in Reproductive and Developmental Toxicological Research\*

### Introduction

Pharmacokinetics is the study of the absorption, distribution, metabolism, and excretion of a chemical agent and a description of these events in mathematical terms. The objective of pharmacokinetics is to define the concentration of the chemical and/or its metabolites in the various compartments at any given time during uptake, distribution, metabolism, and excretion. This information can be used to relate the concentration in critical tissues (i.e., adult reproductive organs, gametes, conceptus, and neonate) to the toxic event (pharmacodynamics) and to help in defining cause and effect (Figure 2). The ultimate goal, however, would be to use this

\*Marjorie G. Horning, Chairperson (Baylor College of Medicine, Houston, TX); Anne B. Brown, IRLG Contact (US Department of Agriculture, Washington, DC); Larry Andrews (ARCO Chemical, New Town Square, PA); Charles E. Becker (University of California, San Francisco, CA); David A. Blake (The Johns Hopkins University, Baltimore, MD); James R. Gillette (National Heart, Lung and Blood Institute/NIH, Bethesda, MD); Jim T. Hill (Chemical Specialties Manufacturers Association, Washington, DC); B. A. Schwetz (DOW Chemical USA, Midland, MI); William Slikker, Jr. (National Center for Toxicological Research, Jefferson, AR); Rochelle Wolkowski-Tyl (Research Triangle Institute, Research Triangle Park, NC); John F. Young (National Center for Toxicological Research, Jefferson, AR).

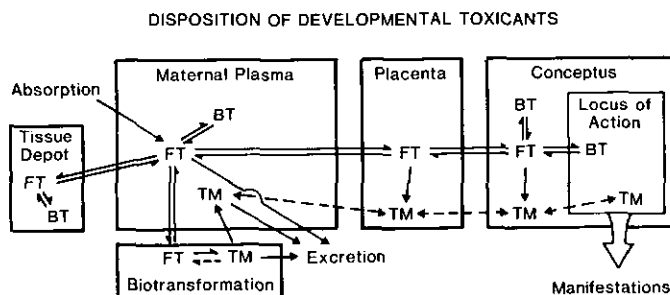


FIGURE 2. Representation of the relation between pharmacokinetics and pharmacodynamics. FT = free toxicant; BT = bound toxicant and TM = toxicant metabolite.

information to predict human reproductive and developmental toxicity. The limiting factor in the use of pharmacokinetic data to predict human reproductive and developmental toxicity is not the state of the art of pharmacokinetics, but rather the ethical and technical considerations which limit the application of the tools of pharmacokinetics to the human situation. Nevertheless, pharmacokinetic data derived from test animal systems can be used, in most instances, to infer from plasma data the limits between which the concentrations of the unbound chemical and its metabolites are likely to occur in animals exhibiting the toxic event.

The discussion of specific issues in the use of pharmacokinetics is limited here to its application to the pregnant female and developing conceptus. Several basic concepts that apply to the maternal-placental-conceptus unit, maternal-neonatal unit and to model animal systems are summarized in the following statements. Figure 3 may aid in its representation of the various compartments involved and their relationships. The first five concepts (a through e) are concerned with parameters that affect concentration of the unbound drug in the maternal-fetal-placental unit. The last three concepts (f through h) comment on dose-response relationships. The importance of a dose-related toxicological response should be recognized. If a toxic response is not related to the quantity of parent compound or one of its metabolites, the validity of the study is open to question.

(a) If a chemical is transferred to the conceptus by simple diffusion and is not metabolized significantly by

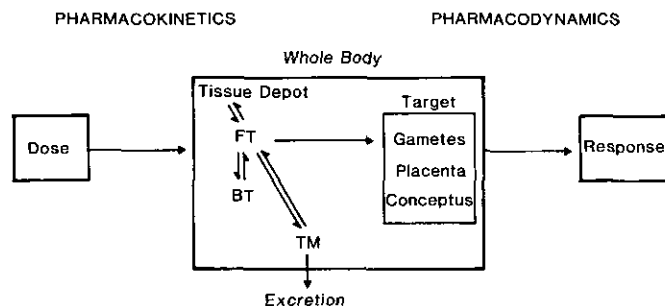


FIGURE 3. Representation of the compartments of the maternal-placental-conceptus unit and the pharmacokinetic parameters that influence the disposition of developmental toxicants.

the conceptus or the placenta or excreted by the fetal kidney, then the pharmacokinetic parameters and the concentration of unbound chemical in maternal blood are useful in predicting the upper limits of unbound chemical in the conceptus (73).

(b) If the chemical is slowly diffusible, the maximum concentration of the chemical in maternal blood may occur much earlier than the maximum concentration in the conceptus. Maternal blood levels, however, can be used to predict the area under the curve (AUC) and thus the average concentration of the unbound chemical in fetal tissue if the chemical is not metabolized by the conceptus or the placenta or excreted by the fetal kidney (73). The maternal blood level would provide an upper limit to the concentration in fetal tissue.

(c) If the chemical is rapidly metabolized by the conceptus or the placenta relative to the rate of diffusion across the placenta, then it is not possible to predict concentrations of the chemical in the conceptus solely from measurements of maternal blood levels (73-76).

(d) If there is active transport of the chemical into, or out of the conceptus, maternal pharmacokinetics alone cannot be used to predict the concentration of unbound chemical in the conceptus (73).

(e) It must be remembered that not all toxic effects on the conceptus are caused solely by a chemical and/or its metabolites within the conceptus. Some effects can be mediated at sites external to the conceptus (for example, in the placenta or in the maternal organism). In such cases, measurement of the chemical or its metabolites in the conceptus would not provide relevant information for predicting risk to the conceptus (77,78).

(f) Pharmacokinetic data spanning the range of doses used in animal (and human) studies facilitate interpretation and use of dose-response curves generated by the toxicological data.

(g) Knowledge of pharmacokinetic parameters in two test species facilitates the selection of dosage regimens for the investigation of species susceptibility at equivalent exposure (36,79,80).

(h) The pharmacokinetics of single dose and repeated dose exposure may be quite different.

Single-dose kinetics are more dependent on the structure of the placenta because diffusional rate constants are more important in limiting concentration of the chemical within the conceptus after single doses than after repeated doses. Nevertheless, as long as the parent compound is not actively transported and the transfer is pH-independent, the assumption that after the administration of a single dose the maximum concentration of unbound parent compound in fetal blood is no higher than the maximum concentration of unbound parent compound in maternal blood is valid.

When repeated doses are administered to steady state, it is likely that the average concentration of the unbound parent compound in the blood of the conceptus is virtually independent of the rate of diffusion across the placenta if the compound is not metabolized by the conceptus and not excreted by fetal kidney. The average



concentration of unbound parent compound in the blood of the conceptus will be less than its concentration in maternal blood if its metabolism occurs within the placenta or conceptus or it is excreted by fetal kidney. The magnitude of the decrease depends on the rate of metabolism and excretion relative to the diffusional rate across the placenta.

### Use of Pharmacokinetic Studies in Choosing Test Animal Species

Pharmacokinetic data are not useful for selection of animal species for initial studies of either reproductive or developmental toxicity if the compound under investigation has not been previously tested for reproductive or developmental toxicity. If the compound is known to exhibit reproductive or developmental toxicity in any species, then pharmacokinetic studies are helpful in elucidating the mechanism of toxicity and establishing relevant parameters to be used in making inter- or intraspecies comparisons. For example, from differences in pharmacokinetic parameters between animal species, it is frequently possible to determine whether the toxicity is due to the parent compound or to a metabolite (81-83).

If it can be established that the toxicity is determined solely by the parent compound, then qualitatively differing patterns of metabolism in different test species may be irrelevant; only the rates of elimination and/or clearance of the compound should be relevant. Pharmacokinetics can determine the dosage levels to be employed to maintain equivalent concentrations of the parent compound in the two species. With such an approach the relative sensitivity of the species to the toxicant can be assessed (36,79,80). If the reproductive toxicity is due to a metabolite, pharmacokinetic data can be used to determine which species are likely to be susceptible based on the ability to form the toxic metabolite (84).

If pharmacokinetic and toxicity data exist for one animal species and comparable pharmacokinetic data exist for a second species, then a qualitatively similar toxic response would be predicted for the second species, if the mechanism of action and target tissue susceptibility are the same in the two species.

### Use of Pharmacokinetic Studies to Establish an Appropriate Dosing Regimen

If the effective dose is defined as the dose at which a toxic effect is manifested in the human conceptus, then pharmacokinetics can help establish a dosing regimen that would result in approximately the same concentration of the chemical in the test system as occurs in the human conceptus after a toxic dose. Owing to the probability that there are quantitative differences in incidence rates and severity of response between species, suitable animal models will usually be selected on the basis of dose-response curves for reproductive and

developmental toxicity rather than pharmacokinetic parameters. Pharmacokinetic studies will then be useful only to determine the relative sensitivities between species after a suitable animal model has been found using toxicity as an endpoint. Pharmacokinetics will also provide a basis for the comparison of various regimens and procedures, i.e., route of administration, vehicle and scheduling of doses.

### Pharmacokinetics and Critical (Sensitive) Periods of Development

Each of the periods of development (preimplantation, organogenesis, and histogenesis) is critical for different aspects of maturation and is affected to different extents by substances which adversely affect development. Animal models are being used to identify the susceptible organ system(s) and gestational time(s) of vulnerability. The possibility exists, however, that any agent delivered to the susceptible site at the critical time in development will yield the same or a similar toxic response irrespective of the animal species or agent, if the mechanism is similar. This statement is predicated on the understanding that the most important determinants for teratogenic susceptibility appear to be the genotype of the organism which determines the kind and level of biochemical response available (80), and the gestational stage of the embryo or fetus at the time of insult. There is some specificity of agent relative to the toxic response but there are almost always lesions in organ systems in addition to the "characteristic anomaly" (85). Once the time of vulnerability has been defined, pharmacokinetics can be used to determine the concentration of the toxicant in the animal model at that time. In the human, pharmacokinetics can assist in quantitating the exposure of the conceptus in specific instances of toxicity. Pharmacokinetics can define the similarities and differences in exposure of the conceptus to a toxicant during critical periods of development in both test animals and humans.

Although studies are being carried out on the pharmacokinetics of drugs in human concepti and newborns (86-91) and in animal models (76,84,92,93), very little information is available about the pharmacokinetics of chemicals in the human fetus during early pregnancy. Furthermore, this information is very difficult to obtain. Pharmacokinetic data can be obtained from the non-pregnant woman, but these data may not be completely applicable. Nevertheless, it should be possible to obtain useful pharmacokinetic data during the first trimester by studying pregnant women on chronic drug therapy (e.g., anticonvulsant, antihypertensive, and antiarthritic drugs). With these data and data from a test animal system, it may be possible to estimate pharmacokinetic parameters for at least these chemicals in the human conceptus during early pregnancy.

If the relationship between maternal and fetal pharmacokinetics is known in one animal system, then knowledge of maternal pharmacokinetics and fetal met-

abolic capacity in another test system should permit reasonable estimates of the pharmacokinetics in the fetus, and the relative risk of a toxicological event.

### Choice of Endpoints (Targets) for Evaluation of Exposure

Ideally one would like to know the correlation between the exposure at the molecular target and the time of initiation of the lesion. It is not possible at this time to establish the concentration of the toxicant at the molecular target by direct measurements. Thus, indirect approaches are necessary, such as measurements of concentrations in specific organs/tissues, embryo/fetal blood, whole embryo/fetus, or maternal blood levels; e.g., by autoradiography or analytical chemical techniques (HPLC, GC-MS) (94). In many cases, it may not be necessary to measure levels in the conceptus because maternal pharmacokinetics can often provide boundaries of the expected concentrations within the conceptus.

Even if the initial biochemical and/or morphological event at the time of insult leading to the ultimate manifestation at term has not been established, pharmacokinetics is useful in evaluating dose-response curves; that is, the relationship of the incidence or severity of the toxic response to the dose administered to the maternal organism and the maternal and/or fetal concentration of the toxicant.

If the mechanism of the toxic event has been established in an animal model, pharmacokinetics would allow the selection of a dosing regimen and sampling intervals that would provide equivalent exposure in additional species. With this information, it should be possible to establish species differences in threshold concentrations and sensitivities. A finding of negligible species differences increases confidence in making extrapolations from animal to man.

### Prediction of Thresholds

There are three kinds of thresholds: metabolic, statistical, and "real." The metabolic threshold can be defined as the concentration (dose) range at which there is a change in the pharmacokinetic parameters (95,96). The statistical threshold may be defined as the dose level below which no adverse effects have been detected within the confines of the experimental design. This is a pragmatic definition which may be dependent on sample size. The "real" threshold can be defined as a dose below which an adverse event would never occur. The "real" threshold cannot be determined in teratological or toxicological experiments at this time. When the molecular mechanism of the toxic event is understood, it may be possible to determine whether a "real" threshold exists and pharmacokinetics may be useful in estimating the threshold.

Pharmacokinetics can identify the existence of a metabolic threshold. In those cases where the metabolic

threshold is related to a toxicologic event (i.e., above a given concentration of toxicant the metabolic profile changes and toxic effects occur in a test system), pharmacokinetics can help to predict the statistical threshold.

### Future Experimentation

Several approaches are needed to better utilize pharmacokinetics in the interpretation of reproductive and developmental toxicity. These are suggested as an impetus for further research in this area.

- Studies to quantify changes in the pharmacokinetics of toxicants throughout gestation. These studies should be designed to describe changes during development in the mother, placenta or conceptus which affect the pharmacokinetics of a chemical (i.e., placental type, active transport, metabolism, blood flow, amniotic fluid disposition, etc.).

- Transplacental pharmacokinetic studies in animal models (e.g., pregnant ewe and monkey) which provide simultaneous, time-course information from fetal and maternal blood.

- Utilization of *in vitro* test systems (derived from the conceptus and placenta) to aid with interpretation of pharmacokinetic data and to relate toxicological events to the concentration of toxicant at the molecular target.

- Utilization of early indicators of chemical effects (pH changes, enzyme activities, heart rate, blood pressure, etc.) to correlate with pharmacokinetic parameters. Even though these early indicators may not necessarily be related to the toxic event, they will provide information concerning the chemical's pharmacodynamic profile.

- Studies in animal models utilizing toxicants for which information is available in pregnant women. These studies can be used to test the validity of pharmacokinetic data gathered from animal models.

- Correlation of pharmacokinetic parameters with teratological endpoints in a single animal. This initial step in prediction must be accomplished in order to identify the pertinent pharmacokinetic and teratological parameters within an animal so that more meaningful interspecies correlation may be attempted.

### Summary

Pharmacokinetics, the application of kinetics to the disposition of chemicals in the body, is a quantification of the absorption, distribution, metabolism, and excretion of chemicals in the intact animal as a function of time. The goal of reproductive toxicological research is to provide a basis for selecting acceptable levels of human exposure to potential toxicants. In the absence of toxicological information, pharmacokinetic data are of limited help in assessing the safety of most chemicals. However, when used in conjunction with data on the toxicology and mechanism of action of chemicals includ-

ing drugs, pharmacokinetic data can be a significant asset in predicting potential risk to humans.

Pharmacokinetics can help in the design of definitive studies after preliminary data are collected. Better characterization of the dose-response curve, more accurate comparison of species sensitivity, accurate definition of the actual dose at the site of action, definition of the rate of transfer of chemicals into and out of the conceptus, and extrapolation between species and from one route of exposure to another can be done more reliably when pharmacokinetic data are available. In addition, those chemicals with saturable elimination kinetics can be identified through pharmacokinetic studies; many such agents with a metabolic threshold have a toxicological profile that correlates with the pharmacokinetic profile. Thus, pharmacokinetic data are important in the prediction of potential developmental and reproductive hazards for humans.

## Workgroup on Risk Assessment in Reproductive and Developmental Toxicology\*

### Introduction

Substantial evidence indicates that agents which have been associated with human reproductive effects can also be associated with reproductive effects in other mammals (8,9,97-102). This suggests that reproductive effects induced in animals may be predictive of reproductive effects in humans. The most substantial evidence for assessing human risk can be derived from human studies with adequate design and sufficient statistical power to detect an effect. However, adequate prevention of adverse effects on human reproduction will only be achieved through prudent use of experimental models.

This report is not intended for the purpose of setting strict criteria for evaluating scientific information which will be taken as evidence (proof) of human reproductive hazard in the absence of human epidemiologic data but rather to review those factors which need to be consid-

ered in reproductive risk assessment for environmental agents. The greater the weight of evidence over a number of factors, the greater the likelihood of a human effect. These factors include but are not limited to: internal consistency of results, reproducibility of results in the same species as well as in the number of species adequately studied with positive results, demonstrated dose-response relationships and routes of exposure, similarities in molecular structure to other known reproductive health hazards, similarities in metabolism and kinetics in the test species to that of humans (when this is known), presence of reproductive effects in the absence of other overt parental toxicity. When human data are available, concordance between experimental studies and human data should also be considered when assessing the evidence, taking into consideration relative exposure levels, power of studies to detect a positive result, and type of endpoints measured.

### Internal Consistency

Internal consistency refers to the need to distinguish biologically significant events from random statistical fluctuations and experimental error. This differs from external consistency, which refers to the need to compare the results among different experimental groups, strains, species, and other factors. In general, the more consistent the results internally, the more qualitative weight may be placed on evidence from the study. Random, statistically significant results are possible in studies due to multiple statistical tests, each employing a small level of significance (e.g., 0.05). For example, a single biological parameter such as a survival index has a probability between 0.15 and 0.46 of occurring at random as a statistically significant event in a three-generation reproduction study (102). "When large numbers of biological parameters are compared in a multi-generation study, the probability of finding at least one statistical false positive is almost a certainty" (103).

The cautious sorting of true effects from statistical artifacts requires expert judgment in line with state of the art thinking in the field. Such judgment entails considerations of biological plausibility, statistical methodology used, a knowledge of historical controls, and evaluation of the conduct of the study.

### Reproducibility of Results in the Same Strain/Species and in Other Strains and Species

Replication of results for a given strain lends considerable weight to a causal hypothesis of effect. When a reproductive effect is found in multiple studies for several species, the evidence is even stronger. A problem in interpretation occurs, however, when results of studies are disparate.

Qualitative evaluation of the degree to which two or more studies are similar with respect to design and

\*Patricia A. Buffler, Chairperson (University of Texas School of Public Health, Houston, TX); Charles Anello, IRLG Contact ((Food and Drug Administration, Rockville, MD); Jerry Chandler (National Institute for Occupational Safety and Health, Rockville, MD); F. W. Carlborg (Consultant, St. Charles, IL); Vernon Chinchilli (Medical College of Virginia, Richmond, VA); Milton Freifeld (Chemical Manufacturers Association, Washington, DC); David Gaylor (National Center for Toxicological Research, Jefferson, AR); James W. Hanson (University of Iowa Hospital and Clinic, Iowa City, IA); Carol J. Hogue (University of Arkansas for Medical Sciences, Little Rock, AR); Rosann Apfeldorf Hurwitz (Occupational Safety and Health Administration, Washington, DC); Jane F. Robens (US Department of Agriculture, Beltsville, MD); Carol Sakai (US Environmental Protection Agency, Washington, DC); Lowell E. Sever (Battelle Northwest Laboratories, Richland, WA); Theodora Tsongas (Occupational Safety and Health Administration, Washington, DC); John Van Ryzin (Columbia University School of Public Health, New York, NY); John R. Wheeler (Standard Oil Company, Chicago, IL).

analysis should be applied to every aspect of the protocols. As an example of the differences in results which can occur when only the timing of effect measurement is varied, Snow and Tam (104) found different effects on fetal growth among mice exposed to mitomycin-C on the same day of gestation, depending on whether sacrifice occurred at 7.5 days, 8.5 days, 10.5 days, 13.5 days, or at birth.

Interspecies and interstrain variations in susceptibility to potential teratogens are well-documented in the literature. Because the causes of interspecies variation may be due to one or more factors, a positive test in one species cannot be regarded as conclusive in determining a given degree of human risk. Such factors include physiological and biochemical differences in maternal pharmacokinetics, differences in placental transfer rates, differences in susceptibility to chemical interactions at molecular, tissue, and organ levels, differences in background incidence of reproductive effects, and differences in the gestational sequence of development. There is no evidence that any particular species or strain more consistently predicts human susceptibility to animal teratogens than any other species or strain (8,97). Hence, it is not possible to specify the number or nature of species or strains that are optimal for human risk assessment. To increase comparability of studies, inherent differences among strains and species should be accounted for to the extent possible by assuring adequate sample sizes for each study (105). The possibility that a human effect could be indicated from animal tests should be further evaluated through considerations of relative exposure levels between the test animal and man, and through considerations of comparative metabolism and pharmacokinetics.

### Dose-Response Relationships

Several potential types of dose-response relationships exist for reproductive effects, depending on the risk under investigation. For example, with teratogenesis, timing of dose as well as level of dose affect response levels (47). As a consequence of the prenatal development of the ova, chronic low-dose exposure is of special potential concern for females. Little is known about dose-response relationships for agents which affect the postnatal development of sexual maturity, for example, for agents that could delay the onset of puberty. In view of the fundamental differences in mechanisms for different reproductive effects, differences in timing of exposure, the paucity of both theoretical quantitative models and experimental evidence, most dose-response data should be interpreted on a case-by-case basis. Interpretation of a particular apparent dose-response curve must include the nature of the effects when considering extrapolation from one species to another; this is especially important when effects are known or suspected to vary across species.

### Structure-Activity Relationships

The observation of reproductive effects associated with some chemical agents raises concern that structural analogs of these compounds may also be active. Notable examples that have been reported are the antifertility effects of various alkylating agents such as dialkylsulfates and epoxides. In contrast with mutagens, the structure-activity relationships for teratogens remains largely unexplored. Nevertheless, structure-activity considerations should be weighed when considering the risk of a particular chemical to human reproduction or development.

### Overt Maternal Toxicity

In mammalian teratogenesis experiments, overt maternal toxicity may occur only at higher dose levels than those which may induce teratogenic effects, or maternal toxicity may occur at dose levels below fetotoxic levels. Three classes of agents can be distinguished based upon our present knowledge: embryo/fetotoxins which are active in the absence of maternal toxicity (e.g., thalidomide, diethylstilbestrol, ionizing radiation); embryo/fetotoxins which elicit overt maternal toxicity (e.g., aminopterin, methylmercury, and polychlorinated biphenyl); and embryo/fetotoxins active at concentrations which induce maternal physiological changes or stresses (e.g., cigarette smoking, steroidal hormone, and alcohol consumption).

Since many agents do not alter embryo/fetal development without exerting some effect on the maternal support system, it is important to thoroughly evaluate the teratogenic potential of agents to which exposure may occur at or near maternally toxic levels.

### Epidemiologic Considerations

Nonexperimental studies of human health effects are based on observations which attempt to relate exposures to health outcomes. Three major issues are involved, all of which must be considered in study design and/or analysis. These issues are: the definition and ascertainment of health outcomes of interest; the definition, identification, and quantification of exposures; and the application of appropriate statistical techniques to determine if an association between exposure and outcome occurs more often than would be expected by chance, and to determine the strength of that association. The following discussion relates to these issues, in the context of reproductive toxicity risk assessment.

There are several considerations in using epidemiologic methods for evaluating agents suspected of inducing adverse human reproductive and/or teratogenic outcomes. These considerations not only apply to the area of reproductive toxicity risk assessment but generally to the use of epidemiologic techniques.

Epidemiology relies on observations from available population groups. While it is often desirable to obtain data for reproductive and teratologic risk assessment from humans, factors in data collection, analysis, and study design often limit the inferences that may be drawn. The "exposed" or "at risk" groups may not be randomly selected and the investigator may have little control over the exposures of the study population, except in a clinical trial. Bias from potential confounding factors and other sources should be minimized through the use of appropriate matching and/or statistical techniques. Unlike experimental investigations, epidemiologic observations are made on individuals exposed during their lives to a variety of hazards. It will not be possible, or plausible, to control for all factors except the one(s) of interest. It is the investigator's obligation to choose the best comparison group(s) possible and note those factors which could bias the outcome. Inappropriate comparison groups will limit the inferences that may be drawn from the analysis.

Objective methods of data collection should be employed in all investigations of reproductive and teratogenic effects. Measures of objectivity may be classified into two categories. The first, validity or accuracy, refers to the extent an observation reflects the actual or true situation. Sensitivity and specificity are two concepts which can be used to describe the validity of a test. The second measure of objectivity is variability or precision, and refers to the consistency or reproducibility of a given observation. It is highly desirable that sufficient data be obtained in a study to describe the objectivity of the methods used, particularly where new methods are being developed or new applications of old methods are being attempted.

#### **Definition and Ascertainment of Outcomes.**

Several sources of data are used to determine the possibility of a reproductive or teratogenic risk. Data from case reports of reproductive or teratogenic effects have often indicated when a potential problem exists. This method relies on the clinician's ability to recognize an unusual occurrence of an event or the association of a health outcome with a particular exposure. Information of this type can generate hypotheses on the nature of reproductive hazards or teratogenic agents.

Vital statistics have the advantage of being routinely collected and readily available. These data are useful for evaluating changes over time. However, certain information may not be recorded, such as measures of infertility, or may not be available for the appropriate geographic area of interest. Certain routinely collected data may not be adequately recorded (e.g., stillbirths).

Data obtained from personal interviews or questionnaires may be subject to reporting bias, memory lapse, or the respondent's unwillingness to provide accurate information. Investigators should select well validated survey instruments and carefully select study and comparison groups to minimize potential problems.

Retrospectively conducted studies such as those using case-control designs are useful in evaluating rare outcomes. Since the events to be studied have already occurred, it is easier to insure that an adequate number of study and comparison subjects are included in the study. Retrospective designs are relatively inexpensive to conduct, but these studies do not prove the existence of a cause and effect relationship.

Prospective studies are not as useful in evaluating rare outcomes, since large numbers of individuals must be followed to insure an adequate number of diseased or exposed individuals in the analysis. Prospective studies are very costly because of the length of time and amount of resources needed to collect data on a large number of individuals. However, they may be used to demonstrate that a cause and effect relationship exists.

**Identification of Exposure.** Poor ascertainment of exposure among study and comparison groups may also limit the inferences from a study. An individual may not recall the amount or duration of exposure to a drug or chemical, or may not be able to specify what levels of a substance were present in the air or work environment. Confirmation of exposures should be attempted from sources other than the individual, such as hospital or company records, or by direct measurement of the work area.

Exposures in general have not been well estimated. Surrogate measures of exposure have included length of residence, length of employment in an exposed occupation, or national estimates of dietary intake, with little or no direct measurement of present exposures. Data on past exposures are rarely available. Timing of exposure in relation to outcome has not been well defined. Difficulties have frequently been encountered in the development of methods for estimation of and/or evaluation of effects of multiple exposure other than by stratification or matching for a few well-known confounders. Finally, the development of an increased utilization of clinical measures is encouraged when they are available and appropriate in the definition of cases or exposed groups.

It is recognized that it is difficult to identify suitable control/comparison/reference groups. Because of this difficulty, appropriate internal or local comparison groups are frequently not included in analytical studies. It should be stressed that every attempt should be made by investigators to identify and include comparison groups, and not to rely solely on the calculation of expected values from national population rates for comparison with observations in exposed groups.

**Statistical Analysis.** The data analysis should be appropriate for the endpoint under evaluation. A single positive result in a well-designed epidemiologic study could indicate a potential hazard. Confirmation of this result in other population groups, if possible, will strengthen the association of the suspected factor with the outcome. Moreover, the association should be bio-

logically plausible. However, a negative or inconclusive study result does not demonstrate the absence of a hazard or the safety of an agent, unless the study was of sufficient statistical power and adequately designed to have been able to detect a difference had one existed. Often, a small number of individuals are exposed, exposures are poorly defined, or the number of observed outcomes is small, thus limiting the statistical ability or power of an investigation to detect a significant excess related to exposure.

Pregnancy outcomes are not independent events and should be adjusted within an analysis. Habitual aborters or parents with a known genetic defect may be excluded. Adjustments for parity and maternal age should be made by investigators in the data analysis or, if possible, in the study design. It is necessary to relate an exposure to an outcome in order to determine risk. Of primary importance are the selection of an appropriate endpoint or endpoints for study, increased precision in the definition and measurement of exposures, and the application of appropriate designs and analytical techniques in the determination of associations and the strength of those associations.

## Epizootic Observations

Epizootic observations from animal populations accidentally exposed to an agent that is suspected of inducing reproductive or teratogenic damage are not usually available in human risk assessment. These observations have the disadvantages of the human studies in that randomization of individuals to "risk" and "nonrisk" groups prior to exposure is not possible. None of the advantages of conventional controlled animal testing are present. Thus, these observations are of interest from an historical perspective as indicators of a potential hazard but yield little further information for human risk assessment.

## Epidemiological Evaluation of Specific Endpoints

**Teratogenesis.** Teratogens share several characteristics. Teratogens usually produce characteristic patterns of abnormality rather than single specific defects. Particular abnormalities of development or growth in individual tissues or organ systems may be caused by many different factors and are therefore not etiologically specific. Variability of effects of teratogenic agents on individuals is common. Few affected individuals display all possible manifestations of damage by a particular agent. This variability can be explained by four basic factors: dosage of agent, timing of exposure, differences in host susceptibility (maternal or fetal), and interactions with other environmental factors. Common indicators of the teratogenic potential of an agent include indications that exposure to the agent is associated with: prenatal-onset growth deficiency, fetal wastage/infertility, abnormalities of morphogenesis, abnormal-

ities of nervous system performance, or carcinogenicity or mutagenicity (though perhaps by different mechanisms).

Past experience has demonstrated that different types of observational or descriptive investigations contribute in specific ways to our understanding of teratogenic hazards to the fetus.

Although case reports, surveillance activities, and descriptive studies can help to generate hypotheses of association between environmental factors and teratogenic outcomes, quantification of risk requires other study designs. Probably the most sensitive and expeditious approach to the confirmation and definition of teratogenic risk in humans is the case-control approach. However, problems of choice of appropriate controls often leave such investigations open to criticism, especially with regard to the specificity and validity of conclusions. Furthermore, direct reliable estimates of attributable risk are difficult without using the prospective cohort approach. Unfortunately, this latter method is a lengthy process which is relatively inflexible and sample-size (power) problems make it difficult to apply strictly for rare events such as birth defects.

Despite these problems, epidemiology has made a significant contribution to human teratogenic risk assessment in many cases by providing confirmatory data on proposed associations, by quantifying risk, and by helping to exclude other risk factors from concern. Nevertheless, new approaches are needed which emphasize biological relationships and which make more efficient use of small sets of human data.

One new approach involves careful clinical pattern analysis and statistical investigation of mechanisms and inter-relationships of outcome variables through a technique called path analysis (106). Since the biological mechanisms underlying the development of clinical patterns of abnormality are presumably interrelated in specific definable ways, these relationships result in biological pathways and developmental hierarchies. Statistical analysis of birth defects on individuals and populations should take this into account and should help to confirm or revise these proposed interactions between outcome variables. This approach may solve many of the design and epidemiologic problems outlined above and may help to define priorities and needs for future human and animal research.

**Spontaneous Abortion.** Spontaneous abortion has been suggested as one of the most useful endpoints for the evaluation of reproductive risks (107). An important aspect of abortion as an endpoint for study relates to the fact that it may be used to identify several different mechanisms leading to early reproductive wastage (108). For example, Stein et al. (109) have presented a model for the relationship between spontaneous abortion and rates of occurrence of congenital malformations.

One of the widely cited advantages of using spontaneous abortions for risk assessment is the frequency with which they occur (107). Because of their frequency,



the power of abortion surveillance or cohort studies to detect an effect of an environmental (nongenetic) factor in reproductive outcome is much greater than for other adverse pregnancy outcomes, such as congenital malformations (110,111).

Currently there are several data sets on spontaneous abortions that are being used for epidemiologic studies. These include the Columbia University Study (112) and the Kaiser Permanente Birth Defects Study (113). There are several aspects of spontaneous abortions that must be considered in evaluating their usefulness for studies of reproductive toxicity. One of these is the critical question of case ascertainment. Unlike congenital malformations, for example, spontaneous abortions may not be readily identified through existing data sources such as vital or hospital records. Hospital based studies can be carried out, such as those at Columbia University, but it is not usually the case that a representative series can be assembled. Case series will usually be assembled through questionnaire studies of reproductive outcome. This method optimally involves some method of validating outcomes.

Abortion surveillance programs offer an important opportunity to monitor the effects of environmental factors. It is important, however, that protocols clearly define what is to constitute a case. In addition, the groups under surveillance must include nonexposed individuals. Appropriate population based data, with which surveillance data can be compared, are limited and estimates of the frequency of early reproductive wastage vary widely (114,115).

It was noted that abortions can represent the outcome of several different mechanisms. These include the action of embryo-lethal agents, chromosomal abnormalities, and structural defects. Optimally, the aborted product of conception will be examined for morphological abnormalities and karyotyped.

Approximately 50% of aborted fetuses are karyotypically abnormal (116). In addition, Hook (117) has suggested that changes in the relative frequencies of aneuploidy and structural rearrangements may relate to the action of environmental factors.

It is extremely important that appropriate control groups be selected in epidemiologic studies of spontaneous abortion. For example, much of the criticism of the Alsea Study (118) revolved around the control groups selected.

Since several common factors, such as alcohol consumption and smoking, have been suggested to increase early reproductive wastage (112), it is important that those factors be considered in data analysis. Methods to control for confounding should be employed (119).

Of importance in relating spontaneous abortions to environmental factors is the timing of exposure. Several instances can be found in the epidemiologic literature where the timing of a purported causal exposure did not occur at a biologically meaningful time (120,121).

Applicable information can be gained from carefully planned prospective studies of noncontracepting

women. Occupationally exposed populations are logical for surveillance but there are a number of legal and sociopolitical considerations. It is clear that appropriate comparison groups must be selected. Detailed information is required on exposure levels and the timing of exposure in relation to gestation age. Although tests are now available to detect pregnancy very early after conception (122), it is probably too early to employ these tests in surveillance studies, since no comparable population based data are available.

**Fertility/Sterility.** Measures of the fertility of a population include crude and age-specific birth rates and general and total fertility rates, all of which examine the number of live births in a certain segment of the population. Fertility has frequently been studied and measured on the basis of other endpoints; for example, the frequency of spontaneous abortion has been used as an indicator of infertility. These measures have usually examined reproductive performance as pregnancy outcome but have generally been insensitive to a lack of conceptions.

Studies of fertility *per se* have been somewhat limited to individual clinical evaluations in the past, and only recently has attention been drawn to potential hazards to fertility in occupational settings. These studies have largely concentrated on the effects of chemical exposures on testicular function, particularly sperm count suppression and abnormal sperm morphology (123-127). These studies have compared the results of semen analysis and other clinical indicators for an exposed group and unexposed comparison groups, or high and low exposure groups. Exposure estimation has varied from the use of direct clinical measures of individuals (i.e., lead in blood or urine) to workplace air monitoring data and/or number of years or months in an exposed occupation.

A second approach to assessing the fertility of occupationally exposed workers has employed data from interviews or self administered questionnaires of exposed workers regarding their and their spouses' reproductive history, primarily the number of live births among wives of male workers (128-130). These studies have compared observed values for number of births to exposed workers to expected values derived from national fertility tables or national birth probabilities. More recent studies (129,130) have calculated and compared standardized fertility ratios (SFR) (observed/expected numbers of births) for exposed and unexposed periods of an employee's reproductive history. Neither of these studies has included a separate unexposed comparison group in the study design. The study of Wong et al. (128) estimated exposures in four plants from historical air monitoring data to workers in broad exposure categories so that dose-related differences were difficult to assess. Levine and coworkers (130) have attempted validation of their methods by examining data from persons exposed to dibromochloropropane who were included in the study of Whorton et al. (124).

The use of national fertility rates or probabilities as

the basis for comparison has been criticized because of the possibility of underestimating expected numbers of births, thereby causing the SFR to be closer to unity in a situation where depressed fertility truly exists (131). It remains to be seen whether this method can be refined by the inclusion in the design of a well characterized comparison group which more closely resembles the occupational group of interest.

These studies have suffered from relatively poor and nonspecific estimates of exposure. The clinically based studies have frequently had difficulties with participation that tended to limit the validity of the study. Semen analysis is a direct and relatively sensitive method for testing fertility. The SFR method is much less sensitive, but is based on information which is easily obtained through questionnaire. Continued development of improvements and refinements in both these methods is desirable.

It has been pointed out (132) that the use of measures of reproductive success rather than failure can avoid difficulties resulting from lost information such as early unrecognized fetal loss, but that the use of age-specific fertility rates is not a good indicator of reproductive problems in low fertility populations. Rather, the period of time required to achieve a particular level of fertility may be a more sensitive indicator of changes in fertility. Further development of this approach may be useful in the development of improved methodologies for studies of fertility.

## Statistical Considerations in Experimental Studies

Three major aspects of sound scientific experimentation that need to be considered in the conduct of teratological and reproductive studies are: control of extraneous factors, randomization of experimental units to treatment groups, and adequate sample size to detect meaningful effects. All of these impinge on the statistical power of the study. The control of extraneous factors is basic to the design of studies in these areas and these factors have been discussed in detail in the various guidelines for studies in reproduction and teratology (22, 133–138). These reports have pointed out a number of critical factors involved in the design, conduct, and interpretation of the tests. This section will discuss other aspects of experimental design particularly as related to power considerations and statistical analysis of the data.

Randomization of experimental units, such as random assignment of animals to treatment and control groups, is essential to eliminate intentional or unintentional biasing of the experimental groups. Proper randomization by using random number tables, for example, minimizes the likelihood that the treatment and control groups will differ substantially with respect to extraneous factors which could influence the experimental results (139).

Determination of an appropriate sample size for an

experiment depends on several factors: the analytic methods to be used, the magnitude of a meaningful effect, and the level of Type I and Type II errors desired by the investigator.

Statistical techniques employed in the analysis will influence the sample size requirements for an experiment. If the endpoint studied is a continuous variable (e.g., birth weight), powerful analytic methods such as analysis of variance (ANOVA) can be used. Generally, if the assumptions required for parametric techniques such as ANOVA are not violated, these techniques are more powerful than nonparametric methods. If the endpoint studied is a categorical variable (e.g., presence or absence of a congenital malformation), the appropriate statistical methods are less powerful because less statistical information is available.

The magnitude of the effect that the experiment is designed to detect and the spontaneous or background rate of the endpoint in question will also influence the sample size requirement for an experiment.

Another consideration in sample size determination is the level of certainty required by the investigator in evaluating the null hypothesis that the exposure is not related to the endpoint. Type I error is the probability of rejecting the null hypothesis when in fact the null hypothesis is correct (false positive). Type II error is the probability of not rejecting the null hypothesis when, in fact, the exposure and outcome are associated (false negative). Obviously both types of error are undesirable and need to be minimized. The relative levels of acceptable Type I and Type II errors should depend on the relative seriousness of falsely implicating an agent that is not a reproductive hazard or exonerating a true reproductive hazard. We should emphasize that sample size determination, involving the components we have just outlined, is far preferable to arbitrary requirements of testing 10, 20, or 50 animals as indicated in various agency guidelines for assessing reproductive toxicity.

In animal teratology studies, there has been a great controversy over whether the entire litter or the individual fetus should be regarded as the experimental unit (140–146). Haseman and Hogan (141) listed several reasons why they felt the litter should be the experimental unit: (a) the pregnant female is randomly assigned to the treatment or control groups; (b) the pregnant female receives the treatment directly; (c) fetuses from the same litter exhibit a "litter effect" and do not respond independently of one another. Assuming that (c) is true, if the fetus is treated as the experimental unit, then the sample size for the statistical test is artificially inflated and the presence of a litter effect would reduce the validity of the test. Endpoints are typically measured according to one of the following scales: continuous, e.g., fetal weight; dichotomous, e.g., one if the fetus has a certain malformation, zero if not; proportional, e.g., the proportion of live fetuses in a litter with a certain malformation.

When individual litter and fetal endpoints are meas-

ured by using a continuous scale, an analysis of variance (ANOVA) procedure with litter nested in the model can be applied for testing litter effects as well as for comparing treatment groups to controls (147). This technique considers the contribution of both between and within litter variance in the analysis of treatment effects. Thus, the individual fetal data are used, and the analysis is conducted by use of the appropriate error term depending on whether or not the litter effect is present.

The dichotomous scale measurement can be used at the individual fetus level or at the litter level. For instance, instead of using the proportion of implantations which result in resorption as a measure of fetolethality, the value one or zero can be assigned to the litter according to whether the female animal has at least one resorption. These dichotomous data for the litter avoid some of the problems encountered with the proportional data in teratology and reproduction experiments; however, their use is not recommended for two basic reasons. First, on using the above example of fetolethality, there is an obvious loss of information; e.g., a female animal with seven resorptions out of ten implantations is assigned the same value as an animal with one resorption out of ten implantations. Second, a more subtle statistical reason is that these dichotomous data do not yield true binomial random variables. With a binomial random variable, it is assumed that the probability of at least one resorption is the same for every female animal in a certain treatment group. However, if one animal has ten implantations and another animal has only two, the first animal has a higher probability of at least one resorption than the second animal, even though they may be the same treatment group. If treatment affects the number of implantations, the dichotomous zero-one values for the litter are not representative of the actual situation, and the statistical analysis based on the comparison of binomial random variables is not justified.

The special type of proportional data encountered in teratology and reproduction studies present some unique statistical problems. For purposes of illustration, assume that the index of interest is the proportion of live fetuses with a certain malformation. For each litter, the measurement is composed of a numerator and a denominator, both of which are random variables. (The denominator is also a random variable because the number of live fetuses within the pregnant animal is not a fixed constant.) Also, the numerator and denominator may be correlated, i.e., when the occurrence of anomalies and death are not independent. Fetal weight presents a special problem as well since it is dependent on litter size. Haseman and Kupper (148) provide an extensive review article on the various statistical procedures available for comparing a treatment group with a control group for this type of proportional data. There is no consensus among statisticians as to what is the best procedure, because no procedure has demonstrated uniform power superiority. With some of the statistical

models discussed below for proportional data, it may be possible to devise a test which would determine the absence or presence of a litter effect. More research is needed in this respect, because the absence of a litter effect means that the fetus could be employed as the experimental unit (and therefore provide a more powerful statistical procedure). Haseman and Kupper (148) discussed four approaches which they consider acceptable for proportional data when a litter effect is present. The generalized binomial models such as the beta-binomial (149) and the correlated binomial (150), are models with extra parameters to allow for the litter effect.

It is possible to rank proportional data for a nonparametric test, as in the Mann-Whitney two-sample test. Gaylor (151) discusses this in detail for teratology and reproduction studies. This avoids some of the distributional problems encountered with the generalized binomial models. However, the Mann-Whitney test does not take into account the varying denominator values. For instance, a response of one out of two is ranked the same as a response of five out of ten, even though the latter response has less inherent variability. Nevertheless, the Mann-Whitney test is an analytical tool which can be very useful for teratology and reproduction studies, because if there is not much variation among the denominator values, it is just as powerful as the generalized binomial models; moreover, it is computationally easy.

Another approach has been to transform proportional values, so that the transformed data are approximately normally distributed, and then to apply ANOVA techniques. Two common transformations are the arc-sine and the Freeman-Tukey binomial arc-sine. This transformation approach, like the Mann-Whitney tests, treats equal proportions in the same manner, regardless of the denominator values.

Gladden (152) applied jackknife methodology for comparing treatment to control groups in teratology and reproduction studies. This technique weights the response according to litter size, and the resultant test statistic has an approximate T-distribution. However, in a small computer simulation, Gladden (152) was not able to show that the jackknife test had any power advantage over the Mann-Whitney test or the transformation approach.

As Haseman and Kupper (148) point out, "It is difficult to recommend unequivocally a particular approach as being superior to all others." No one approach demonstrates a clear power advantage nor provides a better fit of the data. In the case of multiple dose groups, a trend test may be more desirable than individual comparisons to control. For continuous scale data, a contrast for linear trend can be tested within the ANOVA. For proportional data, Jonckheere's test, which is a function of Mann-Whitney statistics, is often utilized. Lin and Haseman (153) have provided a modification of Jonckheere's test in the presence of ties. Some of the other procedures could be modified to yield trend tests, but

this has not been addressed yet in the literature.

A general approach to the simultaneous inference problem (multiple nonindependent endpoints within a given protocol) in teratology and reproduction experiments is very difficult to construct, because quite a variety of designs are employed. An ideal situation for the statistician would be to have the investigator identify a decision rule which defines what constitutes sufficient evidence of teratogenicity or reproductive toxicity. For example, in a three-generation reproduction study, the investigator may label an agent fetotoxic if a significant trend in fetolethality is observed in at least the third generation. Such a decision rule renders the simultaneous inference problem less formidable. Unfortunately, most investigators are not comfortable in relying solely on a decision rule before deciding on the toxicity of a particular agent. For some of the basic designs in teratology and reproduction studies, the consequences of specific decision rules need to be explored. With respect to carcinogenicity experiments, some results have been published by Fears et al. (154) and by Gart et al. (155) in which Bonferroni's inequality is used in conjunction with Fisher's exact test. A similar approach for a one-generation, multiple-dose, teratogenicity study could be adapted by utilizing some of the aforementioned statistical tests and a multiple comparison procedure (such as Bonferroni's inequality). The presence of multiple generations and reproductive endpoints further confounds the problem to the point that the strict use of multiple comparisons would be detrimental to the overall statistical power associated with the experiment.

In the evaluation of teratologic or reproductive effects one must separate the real biologic effects from statistical artifacts. The significance of the biologic effects will depend on the endpoint. For example, a shift in minor skeletal variants such as delayed sternebral ossification would not have the same biologic significance as a cardiac malformation. The former example is a transient effect with no functional consequence, while the latter is a permanent effect and has functional implications which may be incompatible with life. Some endpoints such as skeletal variants must be evaluated on the basis of spontaneous background incidence for the specific strain of test animals.

## Quantitative Risk Assessment

Quantitative risk assessment attempts to relate mathematically risk to exposure and is necessary in order to provide estimates or bounds on potential human teratogenic or reproductive risks. Two major issues involved with risk assessment include those concerned with extrapolation from high to low doses and those concerned with interspecies conversion. The first of these issues is partially a statistical one which can be addressed with available mathematical models, whereas the latter, interspecies conversion, is primarily a problem of biological interpretation.

High- to low-dose extrapolation is often necessary because experimental studies typically must use high doses, which are well above human exposure levels, in order to detect potential toxic effects using limited numbers of experimental animals (or units). Two ways of developing such estimates are the use of safety factors and the use of mathematical models for extrapolation.

Several limitations exist in the use of the safety factor approach. For example, it is not known how large a safety factor should be since relationships have not been established between the magnitude of the safety factors applied and the desired risk levels (156). A common shortcoming associated with the use of safety factors is the application of safety factors to "no observed effect levels," a process which results in some inconsistencies. The apparent "no observed effect level" is dependent upon the sample size in the dose group, the locations of the dose groups on the dose-response curve and the background level for the particular endpoint evaluated. In one case, the true response rate at the "no observed effect level" may be 0.1%, in another case 20.0%, etc. It must be emphasized that the "no observed effect level" is not the same as an actual no-effect level (threshold) for the entire population. For example, in a teratology study with no background response rate, a treated group of 20 animals showing no adverse response would have a upper 95% confidence limit of 14% on the response rate. Thus, an apparent no observable effect is not proof of the existence of a threshold but only demonstrates a response rate below 14% with 95% confidence in this case. Safety factors applied to data generated from animal experiments for the high- to low-dose extrapolation, also may not conclude considerations concerning interspecies conversion.

There are also limitations inherent in high- to low-dose extrapolations using mathematical models. Statistical variability of the data within the experimental range results in some uncertainty. Outside the experimental dose range additional uncertainty exists because of the unknown shape of the dose-response curve.

Several models are available which generally fit the data within the experimental dose range. However, when extrapolating beyond the experimental dose range, full reliance on the model will often lead to larger differences in estimates of the dose associated with a specific risk, varying in some cases as much as several orders of magnitude (157). Thus, the resulting estimate depends on the model selected. The correct model is (almost) always unknown. However, it is possible to establish conservative upper bounds on risk at low doses in order to establish the corresponding allowable dose limits.

## Available Extrapolation Methods for Different Types of Data

Depending on the type of data available or endpoints selected, different methods of extrapolation are appropriate. Three types of data are commonly observed in

teratology/reproductive studies: continuous data, count data, and quantal (binary) data.

**Continuous Data.** Continuous data are those which are measured on a continuous or finely partitioned scale. Examples of these are birth weights of fetuses, hormone levels, or certain behavioral measurements. For these types of data, many statistical procedures exist to estimate response and confidence intervals for estimates in the experimental dose range. Simple or weighted regression techniques for dose-response analysis are appropriate. If the continuous experimental response is within the range of human levels of interest, then there is no extrapolation problem. This may be the case for some endpoints such as fetal weight reduction. If, however, the experimental responses are outside the range of interest for human levels, then extrapolation outside the range becomes a problem. Best estimates of response levels outside the observed range are sensitive to modeling assumptions. However, in most regression models confidence limits can be assigned for extrapolation outside the experimental range. Typically, if one is far outside this range, the confidence intervals will be so broad that they allow for considerable spread in possible estimates.

**Count Data.** Count data are measured on an integer scale such as litter size. Often ratios of count data are formed, such as the proportion of anomalies per litter or proportion of resorbed fetuses and/or dead fetuses per litter. Ratios of count data or transformations of these data can be analyzed by methods similar to those used for continuous data both within and outside the experimental data range (148). Caution, however, should be exercised when using ratios in the presence of reduced litter sizes if there is a dependent relationship between anomalies and fetal death (158). In addition, linear extrapolation to zero below the experimental data range, after accounting for spontaneous background rates, provide upper bounds on these ratios at low doses assuming upward curvature of the low-dose response (159). If the data do not demonstrate upward curvature, other extrapolation procedures may be required.

For count data such as litter size, techniques of modeling discrete distributions whose parameters depend on dose may be used (148). Another alternative is to use transformations of count data which are amenable to continuous techniques.

**Quantal (Binary) Data.** For quantal data there exist a variety of dose-response models in the literature based on modeling concepts from both the tolerance distribution viewpoint and stochastic mechanism viewpoint (160). Many of these might be appropriate for various binary endpoints in teratology or reproduction studies. Furthermore, for any experimental data set, more than one of these models will fit the observed data adequately. Thus, when it comes to estimating the risk, i.e., the probability of the adverse response, the issue of extrapolation will again arise for response levels below the observed levels. Again, the estimates outside

the experimental dose range will be highly model dependent, therefore preventing accurate estimates of dose levels corresponding to prescribed low risk levels. However, if the dose-response curves upward in the low dose range, linear extrapolation below the experimental range will provide upper bounds on-risk (159). Again, if the data do not demonstrate upward curvature, other extrapolation procedures may be required.

## Experimental Design

Experimental animal studies, designed for screening or detection of toxicity may have limited use for extrapolation. In designing a study for extrapolation, it is desirable to have a minimum of at least three dose levels covering an adequate range of responses above the background rate. Sample sizes at each dose level depend on whether the endpoint is based on litters or fetuses and the desired level of precision for the estimates of response.

## Interspecies Conversion

Up to this point, only the estimation of risk within the experimental study population has been addressed. Interspecies conversion has typically been ignored or addressed by the use of safety factors. The determination of the size of safety factors for interspecies conversion is not a mathematical problem, but rather one involving biological interpretation. Guidelines for such interpretation are discussed elsewhere (see reports on endpoints of teratogenicity and reproductive toxicity).

## Epidemiology

When dose-response information is available from an epidemiological study, quantitative estimates of risk may be obtained in a manner similar to animal studies. In this case the extrapolation problem is minimized as the exposure levels in the study group generally will be closer to the exposure levels in the population of interest. Exposure levels may not be well known, but best estimates of exposure may be used. Multiple exposures and confounding variables must be identified and adjustments made for their effects in either the design or the analysis. Bias resulting from the selection of cases and controls may limit the population of inference in epidemiologic studies.

## Clinical Trials

Clinical trials provide data from controlled randomized studies in human populations. As such, confounding variables may be less of a problem with data from clinical trials; however, the population of inference may be rather restricted. Data from clinical trials may be used for quantitative risk estimates when dose-response information exists.

## Summary

Numerous uncertainties make it difficult to evaluate the reliability of animal experiments as predictors of human reproductive and developmental toxicity. Human data on the various agents studied are of varying quality and quantity, and it is possible that as additional data become available, earlier conclusions may be modified. In addition, there are marked differences in the response of different animal species to chemical agents and there is no clear agreement as to which species is likely to be the most appropriate or how many species should be evaluated.

Reproductive toxicity in humans is not observed for many chemicals which affect reproduction in animals. With the limited available information for these exposures, it cannot be determined if the negative or inconclusive results in human studies are due to the lack of human response to these agents, low susceptibility in humans, lack of sufficient exposure, or inadequacies of the studies.

While these uncertainties impede our ability to evaluate the reliability of animal studies as predictors of human risk, for many agents, results from animal studies may be the only data available.

Factors to be considered when reviewing animal and human data for quantitative and qualitative risk assessment of human reproductive hazards are discussed. Epidemiologic methods for collection of data and statistical techniques for the analysis of data are outlined and certain guidelines provided in the hope that the limited available data may be used effectively to assist the prediction of the potential for risk to the human population and to guide the conduct of needed additional animal and human studies.

## REFERENCES

- Food and Drug Administration. Reproductive Toxicity Risk Assessment Task Group: outline of work plan and request for comments. Fed. Reg. 45(188): 63553-63554 (1980).
- Galbraith, W. M., Voytek, P., and Ryon, M. S. (Eds.). Assessment of risks to human reproduction and to development of the human conceptus from exposure to environmental substances. In: *Advances in Modern Environmental Toxicology*, Vol. III. Assessment of Reproductive and Teratogenic Hazards, Princeton Scientific, Princeton, NJ, 1983.
- Dixon, R. L. (Ed.). Proceedings of the NTP/NIEHS Reproductive Toxicology Workgroup, Research Triangle Park, NC, 1980.
- Legator, M. S., Rosenberg, M. J., and Zenick, H. (Eds.). Environmental influence on fertility, pregnancy and development. *Teratogen. Carcinogen. Mutagen.* 4(1): 1-169 (1984).
- U.S. Environmental Protection Agency. Proposed guidelines for the health assessment of suspect developmental toxicants and request for comments. Fed. Reg. 49(227): 46323-46331 (1984).
- Kalter, H. The natural elimination of sporadically malformed mouse embryos. *Teratology* 22: 201-205 (1980).
- Johnson, E. M., Nelson, M. M., and Monie, I. W. Effects of transitory pteroylglutamic acid (PGA) deficiency on embryonic and placental development in the rat. *Anat. Rec.* 146: 215-224 (1963).
- Schardein, J. L. Teratogenic risk assessment. In: *Issues and Reviews in Teratology* (H. Kalter, Ed.), Plenum Press, New York, 1983, Vol. 1, pp. 181-214.
- Nishimura, H., and Shiota, K. Summary of comparative embryology and teratology. In: *Handbook of Teratology* (J. G. Wilson and F. C. Fraser, Eds.), Plenum Press, New York, 1977, pp. 119-154.
- Palmer, A. K. Assessment of current test procedures. *Environ. Health Perspect.* 18: 97-104 (1976).
- Palmer, A. K. Regulatory requirements for reproductive toxicology: theory and practice. In: *Developmental Toxicology* (C. A. Kimmel and J. Buelke-Sam, Eds.), Raven Press, New York, 1981, pp. 259-287.
- Johnson, E. M. Screening for teratogenic potential: are we asking the proper question? *Teratology* 21: 259 (1980).
- Johnson, E. M. Screening for teratogenic hazards: nature of the problems. *Ann. Rev. Pharmacol. Toxicol.* 21: 417-429 (1981).
- Donovan, J. W. Randomized controlled trial of anti-smoking advice in pregnancy. *Brit. J. Prevent. Soc. Med.* 31: 6-12 (1977).
- Rosett, H. L., Ouellette, E. M., Weiner, L., and Owens, E. Therapy of heavy drinking during pregnancy. *Obstet. Gynecol.* 51: 41-46 (1978).
- Beck, F., and Lloyd, J. B. Comparative placental transfer. In: *Handbook of Teratology*, Vol. 3 (J. G. Wilson and F. C. Fraser, Eds.), Plenum Press, New York, 1977, pp. 155-186.
- Palmer, A. K. Incidence of sporadic malformations, anomalies and variations in random bred laboratory animals. In: *Methods in Prenatal Toxicology* (D. Neubert, H.-J. Merker, and T. E. Kwasigroch, Eds.), Georg Thieme Publisher, Stuttgart, 1977, pp. 52-71.
- Khera, K. S. Common fetal aberrations and their teratologic significance: A review. *Fundam. Appl. Toxicol.* 1: 13-18 (1981).
- Shrader, R. E., Joy, N., and Zerman, F. J. Fetal ossification site development in induced hypothyroidism in rats. *Nutr. Reports Intl.* 20: 403-410 (1979).
- Aliverti, V., Bonanomi, L., Giavini, E., Leone, V. G., and Mariani, L. The extent of delayed development in teratogenic studies on the rat. *Teratology* 20: 237-242 (1979).
- Kimmel, C. A. A profile of developmental toxicity. In: *Developmental Toxicology* (C. A. Kimmel and J. Buelke-Sam, Eds.), Raven Press, New York, 1981, pp. 321-331.
- Food and Drug Administration, Advisory Committee on Protocols for Safety Evaluations. Panel on reproduction report on reproductive studies in the safety evaluation of food additives and pesticide residues. *Toxicol. Appl. Pharmacol.* 16: 264-296 (1970).
- Christian, M. S. Postnatal alterations of gastrointestinal physiology, hematology, clinical chemistry and other non-CNS parameters. In: *Teratogenesis and Reproductive Toxicology* (E. M. Johnson and D. M. Kochhar, Eds.), Springer-Verlag, Berlin, 1982, pp. 263-286.
- Jelinek, R., and Rychter, Z. Morphogenic systems and the central phenomena of teratology. In: *Advances in the Study of Birth Defects*, 2nd Ed. (T. V. N. Persaud, Ed.), Univ. Park Press, Baltimore, 1979, pp. 41-67.
- Byrd, R. A., Kimmel, C. A., Morris, M. D., Holson, J. F., and Young, J. F. Altered pattern of prenatal toxicity in rats due to different treatment schedules with Mirex. *Toxicol. Appl. Pharmacol.* 60: 213-219 (1981).
- Herbst, A. L., Ulfelder, H., and Poskanzer, D. C. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. *N. Engl. J. Med.* 284: 878-881 (1971).
- Johnson, E. M. A prioritization of biological decision tree for developmental toxicity safety evaluations. *J. Am. Coll. Toxicol.* 3: 141-147 (1984).
- Kochhar, D. M., Penner, J. D., and McDay, J. A. Limb development in mouse embryos. II. Reduction defects, cytotoxicity and inhibition of DNA synthesis produced by cytosine arabinoside. *Teratology* 18: 71-92 (1978).
- Pratt, R. M., and King, C. T. G. Inhibition of collagen cross-linking associated with beta-aminopropionitrile-induced cleft palate in the rat. *Dev. Biol.* 27: 322-328 (1972).
- Wilson, J. G. *Environment and Birth Defects*. Academic Press, New York and London, 1973.



31. Scott, W. J., Ritter, E. J., and Wilson, J. G. DNA synthesis inhibition and cell death associated with hydroxyurea teratogenesis in rat embryos. *Dev. Biol.* 26: 306-315 (1971).
32. Butcher, R. E., Scott, W. J., Kazmaier, K., and Ritter, E. J. Postnatal effects in rats of prenatal treatment with hydroxyurea. *Teratology* 7: 161-166 (1973).
33. Ritter, E. J., Scott, W. J., and Wilson, J. G. Teratogenesis and inhibition of DNA synthesis induced in rat embryos by cytosine arabinoside. *Teratology* 4: 7-13 (1971).
34. Keeler, R. F. Teratogenic compounds of *Veratrum californicum* (Durant). *Teratology* 3: 175-180 (1970).
35. Chernoff, G. F. The fetal alcohol syndrome in mice: maternal variables. *Teratology* 22: 71-75 (1980).
36. Wilson, J. G., Ritter, E. J., Scott, W. J., and Fradkin, R. Comparative distribution and embryotoxicity of acetylsalicylic acid in pregnant monkeys and rats. *Toxicol. Appl. Pharmacol.* 41: 67-68 (1977).
37. Beck, F., Lloyd, J. B., and Griffiths, A. Lysosomal enzyme inhibition by trypan blue: a theory of teratogenesis. *Science* 157: 1180-1182 (1967).
38. Goldman, A. S., Katsumata, M., Yaffe, S. J., and Gasser, D. L. Palatal cytosol cortisol-binding protein associated with cleft palate susceptibility and H-2 genotype. *Nature* 265: 643-645 (1977).
39. Salomon, D. S., and Pratt, R. M. Involvement of glucocorticoids in the development of the secondary palate. *Differentiation* 13: 141-154 (1979).
40. Wilson, J. G., Maren, T. H., Takano, K., and Ellison, A. Teratogenic action of carbonic anhydrase inhibitors in the rat. *Teratology* 1: 51-60 (1968).
41. Schreiner, C. M., Hirsch, K. S., and Scott, W. J. Carbonic anhydrase distribution in rodent embryos and its relationship to acetazolamide teratogenesis. *J. Histochem. Cytochem.* 29: 1213-1218 (1981).
42. Scott, W. J., Hirsch, K. S., DeSesso, J. M., and Wilson, J. G. Comparative studies on acetazolamide teratogenesis in pregnant rats, rabbits and rhesus monkeys. *Teratology* 24: 37-42 (1981).
43. Ritter, E. J., Scott, W. J., and Wilson, J. G. Embryotoxic interactions between inhibitors of diverse biosynthetic pathways in the rat. *Teratology* 21: 42A (1980).
44. Millicovsky, W. G. Augmentation of hydroxyurea teratogenesis by drug-induced mast cell degranulation. *Anat. Rec.* 193: 625 (1979).
45. Ritter, E. J., Scott, W. J., Wilson, J. G., Mathinos, P. R., and Randall, J. L. Potentiation interactions between caffeine and various teratogenic agents. *Teratology* 25: 95-100 (1982).
46. Chaube, S., and Murphy, L. M. Protective effect of deoxycytidylic acid (CDMP) on hydroxyurea-induced malformations in rats. *Teratology* 7: 79-87 (1973).
47. Wilson, J. G. Current status of teratology. General principles and mechanisms derived from animal studies. In: *Handbook of Teratology*, Vol. 1 (J. G. Wilson and F. C. Fraser, Eds.), Plenum Press, New York, 1977, pp. 47-74.
48. DeSesso, J. M. Amelioration of teratogenesis. I. Modification of hydroxyurea-induced teratogenesis by the antioxidant propyl gallate. *Teratology* 24: 19-35 (1981).
49. Overstreet, J. W. Laboratory tests for human male reproductive risk assessment. *Teratogen. Carcinogen. Mutagen.* 4: 67-82 (1984).
50. Collins, T. F. X. Multigeneration reproduction studies. In: *Handbook of Teratology*, Vol. 4, Research Procedures and Data Analysis (J. G. Wilson and F. C. Fraser, Eds.), Plenum Press, New York, 1978, pp. 191-214.
51. Mattison, D. R. How xenobiotic compounds can destroy oocytes. *Contemp. Ob. Gyn.* 15: 157-169 (1980).
52. Yen, S. S. C., and Jaffee, R. B. *Reproductive Endocrinology*. W. B. Saunders Co., Philadelphia, 1978.
53. Edwards, R. G. *Conception in the Human Female*. Academic Press, New York, 1980.
54. Lipshultz, L. I., and Howard, S. S. *Infertility in the Male*. Churchill Livingstone, New York, 1983.
55. Eroschenko, V., and Palmiter, R. Estrogenicity of kepone in birds and mammals. In: *Estrogens in the Environment* (J. McLachlan, Ed.), Elsevier/North Holland, New York, 1980, p. 305.
56. Kupfer, D., and Bulger, W. H. Estrogenic properties of DDT and its analogs. In: *Estrogens in the Environment* (J. McLachlan, Ed.), Elsevier/North Holland, New York, 1980, p. 239.
57. Clark, J. H., and Peck, E. J., Jr. *Female Sex Steroids: Receptors and Functions*. Springer-Verlag, Berlin, Heidelberg, New York, 1979.
58. MacLusky, N. J., and Naftolin, F. Sexual differentiation of the central nervous system. *Science* 211: 1294 (1981).
59. Mattison, D. R., and Thorgeirsson, S. S. Ovarian arylhydrocarbon hydroxylase activity and primordial oocyte toxicity of polycyclic aromatic hydrocarbons in mice. *Cancer Res.* 39: 3471-3475 (1979).
60. Mattison, D. R., and Thorgeirsson, S. S. Smoking and industrial pollution and their effects on menopause and ovarian cancer. *Lancet* i: 187-188 (1978).
61. Chapman, R. M. Gonadal injury resulting from chemotherapy. *Am. J. Ind. Med.* 4: 149-161 (1983).
62. Teramoto, S., Saito, R., Aoyama, H., and Shirasu, Y. Dominant lethal mutation induced in male rats by 1,2-dibromo-3-chloropropane (DBCP). *Mutat. Res.* 77: 71-78 (1980).
63. Lee, I. P., and Suzuki, K. Induction of unscheduled DNA synthesis in mouse germ cells following 1,2-dibromo-3-chloropropane (DBCP) exposure. *Mutat. Res.* 68: 169-173 (1979).
64. Kluwe, W. M., Gupta, B. N., and Lamb, J. C. IV. The comparative effects of 1,2-dibromo-3-chloropropane (DBCP) and its metabolites, 3-chloro-1,2-propanedioide (epichlorohydrin), 3-chloro-1,2-propanediol (aliphachlorohydrin), and oxalic acid, on the urogenital system of male rats. *Toxicol. Appl. Pharmacol.* 70: 67-86 (1983).
65. Sotomayer, R. E., and Cumming, R. B. Induction of translocations by cyclophosphamide in different germ cell stages in male mice: cytological characterization and transmission. *Mutat. Res.* 27: 375-388 (1975).
66. Pacchierotti, F., Bellincampi, D., and Civitareale, D. Cytogenetic observations, in mouse secondary spermatocytes, on numerical and structural chromosome aberrations induced by cyclophosphamide in various stages of spermatogenesis. *Mutat. Res.* 119: 177-183 (1983).
67. Generoso, W. M., and Russell, W. L. Strain and sex variations in the sensitivity to dominant-lethal induction with ethyl methanesulfonate. *Mutat. Res.* 8: 589-598 (1969).
68. Lee, I. P. Adaptive biochemical repair response toward germ cell DNA damage. *Am. J. Ind. Med.* 4: 135-147 (1983).
69. Hansmann, I. Chromosome aberrations in metaphase II-oocytes, stage sensitivity in mouse oogenesis to amethopterin and cyclophosphamide. *Mutat. Res.* 22: 175-191 (1974).
70. Hardin, B. D. Reproductive toxicity of the glycol ethers. *Toxicology* 27: 91-102 (1983).
71. Foster, P. M. D., Creasy, D. M., Foster, J. R., Thomas, L. V., Cook, M. W., and Gangolli, S. D. Testicular toxicity of ethylene glycol monomethyl and monoethyl ethers in the rat. *Toxicol. Appl. Pharmacol.* 69: 385-399 (1983).
72. Chapin, R. E., and Lamb, J. C. IV. Effects of ethylene glycol monomethyl ether on various parameters of testicular function in the F344 rat. *Environ. Health Perspect.* 57: 219-224 (1984).
73. Gillette, J. R. Chairman's introductory statement: prenatal development of drug metabolism. In: *The Introduction of Drug Metabolism* (R. W. Estabrook and E. Lindenlauf, Eds.) Springer Verlag, Stuttgart-New York, 1979, pp. 501-506.
74. Slikker, W., Jr., Hill, D. E., and Young, J. F. Comparison of the transplacental pharmacokinetics of estradiol-17 $\beta$  and diethylstilbestrol in the sub-human primate. *J. Pharmacol. Exptl. Therap.* 221: 173-182 (1982).
75. Slikker, W., Jr., Bailey, J. R., Newport, G. D., Lipe, G. W., and Hill, D. E. Placental transfer and metabolism of 17 $\alpha$ -ethynylestradiol-17 $\beta$  and estradiol-17 $\beta$  (E<sub>2</sub>) in the rhesus monkey. *J. Pharmacol. Exptl. Therap.* 223: 483-489 (1982).
76. Slikker, W., Jr., Althaus, Z. R., Rowland, J. M., Hendrickx, A. G., and Hill, D. E. Comparison of the metabolism of cortisol

- and triamcinolone acetonide in the early, mid and late gestational age rhesus monkey (*Macaca mulatta*). *Dev. Pharmacol. Ther.* 7: 319-333 (1984).
77. Hurley, L. S., Gowan, J., and Swenerton, H. Teratogenic effects of short-term and transitory zinc deficiency in rats. *Teratology* 4: 199-204 (1971).
  78. Hurley, L. S., and Mutch, P. B. Prenatal and postnatal development after transitory gestational zinc deficiency in rats. *J. Nutr.* 103: 649-657 (1973).
  79. Wilson, J. G., Scott, W. J., Ritter, E. J., and Fradkin, R. Comparative distribution and embryotoxicity of hydroxyurea in pregnant rats and rhesus monkeys. *Teratology* 11: 169-178 (1975).
  80. Shum, S., Jensen, N. M., and Nebert, D. N. The murine *Ah* locus: *in utero* toxicity and teratogenesis associated with genetic differences in benzo[a]pyrene metabolism. *Teratology*, 20: 365-376 (1979).
  81. Harbison, R. D., and Becker, B. A. Effect of phenobarbital and SKF 525A pretreatment on diphenylhydantoin teratogenicity in mice. *J. Pharmacol. Exptl. Therap.* 195: 283-288 (1970).
  82. Harbison, R. D., and Becker, B. A. Effects of phenobarbital or SKF 525A pretreatment on diphenylhydantoin disposition in pregnant mice. *Toxicol. Appl. Pharmacol.* 20: 573-581 (1971).
  83. Harbison, R. D., Mantilla-Planta, B., and Lubin, D. J. Alteration of  $\Delta^9$ -tetrahydrocannabinol-induced teratogenicity by stimulation and inhibition of its metabolism. *J. Pharmacol. Exptl. Therap.* 202: 455-465 (1977).
  84. Gordon, G. B., Spielberg, S. P., Blake, D. A., and Balasubramanian, V. Thalidomide teratogenesis: Evidence for a toxic arene oxide metabolite. *Proc. Natl. Acad. Sci. (U.S.)* 78(4): 2545-2548 (1981).
  85. Wolkowski-Tyl, R. Reproduction and teratogenic events: No more thalidomides. In: *The Pesticide Chemist and Modern Toxicology* (ACS Symp. Series Vol. 160) American Chemical Society, Washington, DC, 1981, pp. 115-155.
  86. Bhat, R., Vidyasagar, D. V., Vadapalli, M., Whalley, C., Fisher, E., Hastreiter, A., and Evans, M. Disposition of indomethacin in preterm infants. *J. Pediatr.* 95: 313-316 (1979).
  87. Thalji, A. A., Carr, I., Yeh, T. F., Raval, D., Luken, J. A., and Pildes, R. S. Pharmacokinetics of intravenously administered indomethacin in premature infants. *J. Pediatr.* 97: 995-1000 (1980).
  88. Vert, P., Blanchetti, G., Marchal, F., Monin, P., and Morselli, P. L. Effectiveness and pharmacokinetics of indomethacin in premature newborns with patent ductus arteriosus. *Europ. J. Clin. Pharmacol.* 18: 83-88 (1980).
  89. Juchau, M. R., Chao, S. T., and Omiecinski, C. J. Drug metabolism of the human fetus. *Clin. Pharmacokinet.* 5: 320-339 (1980).
  90. Yaffe, S. J., Friedman, W. F., Rogers, D., Lang, P., Ragni, M., and Saccar, C. The disposition of indomethacin in preterm babies. *J. Pediatr.* 97: 1001-1006 (1980).
  91. Olanow, C. W., and Finn, A. L. Phenytoin: pharmacokinetics and chemical therapeutics. *Neurosurgery* 8: 112-117 (1981).
  92. Bus, J. S., White, E. L., Tyl, R. W., and Barrow, C. S. Perinatal toxicity and metabolism of *n*-hexane in Fischer-344 rats after inhalation exposure during gestation. *Toxicol. Appl. Pharmacol.* 51: 295-302 (1979).
  93. Midha, K. K., Buttar, H. S., Rowe, M., and Dupuis, I. Metabolism and disposition of trimethadione in pregnant rats. *Epilepsia* 20: 417-423 (1979).
  94. Irons, R. D., Gross, E. A., and White, E. L. Aniline: evidence for an enterogastric cycle in the rat. *Food Cosmet. Toxicol.* 18: 393-397 (1980).
  95. McKenna, M. J., and Zempel, J. A. The dose-dependent metabolism of  $^{14}$ C-methylene chloride following oral administration to rats. *Food Cosmet. Toxicol.* 19: 73-78 (1981).
  96. Davis, D. C., Potter, W. Z., Jollow, D. J., and Mitchell, J. R. Species differences in hepatic glutathione depletion, covalent binding and necrosis after acetaminophen. *Life Sci.* 14: 2099-2109 (1974).
  97. Council on Environmental Quality, Chemical Hazards to Human Reproduction. Clement Associates Inc., Washington, DC, 1981.
  98. Wilson, J. G. Embryotoxicity of drugs in man. In: *Handbook of Teratology*, Vol. 1 (J. G. Wilson and F. C. Fraser, Eds.), Plenum Press, New York, 1977, pp. 309-355.
  99. Fraser, F. C. Relation of animal studies to the problem in man. In: *Handbook of Teratology*, Vol. 1 (J. G. Wilson and F. C. Fraser, Eds.), Plenum Press, New York, 1977, pp. 75-96.
  100. Schardein, J. L. *Drugs as Teratogens*. CRC Press, Cleveland, 1976.
  101. Food and Drug Administration. Caffeine: deletion of GRAS status, proposed declaration that no prior sanction exists, and use on an interim basis pending additional study. Proposed rule. *Fed. Reg.* 45(205): 69816-69838 (1980).
  102. Steele, R. G. D., and Torrie, J. H. *Principles and Procedures of Statistics*, 2nd ed., McGraw-Hill, New York, 1980, pp. 137-148, 188-190.
  103. Rao, K. S., Schwetz, B. A., and Park, C. N. Reproductive toxicity risk assessment of chemicals. *Vet. Human Toxicol.* 23: 167-175 (1981).
  104. Snow, M. H. L., and Tam, P. P. L. Is compensatory growth a complicating factor in mouse teratology? *Nature* 279: 555-557 (1979).
  105. Nelson, C. J., and Holson, J. F. Statistical analysis of teratologic data: problems and advancements. *J. Environ. Pathol. Toxicol.* 2: 187-199 (1978).
  106. Li, C. C. *Population Genetics*. Boxwood Press, Pacific Grove, CA, 1976, pp. 255-289.
  107. Kline, J., Stein, Z., Strobino, B., Susser, M., and Warburton, D. Surveillance of spontaneous abortions. Power in environmental monitoring. *Am. J. Epidemiol.* 106: 345-350 (1977).
  108. Sever, L. E. Reproductive hazards of the workplace. *J. Occup. Med.* 23: 685-689 (1981).
  109. Stein, Z., Susser, M., Warburton, D., Wittes, J., and Kline, J. Spontaneous abortion as a screening device. The effect of fetal survival on the incidence of birth defects. *Am. J. Epidemiol.* 102: 275-290 (1975).
  110. Selevan, S. G. Design considerations in pregnancy outcomes. Studies of occupational populations. *Scand. J. Work Environ. Health* 7 (Suppl. 4): 76-82 (1981).
  111. Hatch, M. C. Spontaneous abortion as an endpoint in environmental research. Proceedings of the Latin American Course in Genetic Toxicology, Mexico City, 1981.
  112. Kline, J., Stein, Z., Susser, M., and Warburton, D. Environmental influences on early reproductive loss in a current New York City study. In: *Human Embryonic and Fetal Death* (I. H. Porter and E. B. Hook, Eds.), Academic Press, New York, 1980, pp. 225-240.
  113. Harlap, S., Shiono, P. H., and Ramcharan, S. A life table of spontaneous abortions and the effects of age, parity and other variables. In: *Human Embryonic and Fetal Death* (I. H. Porter and E. B. Hook, Eds.), Academic Press, New York, 1980, pp. 145-158.
  114. Roberts, C. J., and Lowe, C. R. Where have all the conceptions gone? *Lancet* i: 498-499 (1975).
  115. Shepard, T. H., and Fantel, A. G. Embryonic and early fetal loss. *Clinics Perinatol.* 6: 219-243 (1979).
  116. Boue, J., Boue, A., and Lazar, P. Retrospective and prospective epidemiological studies of 1500 karyotyped spontaneous human abortions. *Teratology* 12: 11-26 (1975).
  117. Hook, E. B. Human germinal mutations: monitoring for environmental effects. *Social Biol.* 22: 105-116 (1979).
  118. Milby, T. H., Husting, E. L., Whorton, M. D., and Larson, S. Potential Health Effects Associated with the Use of Phenoxy Herbicides: A Summary of Recent Scientific Literature. Report for the National Forest Products Association, Environmental Health Associated, Inc., Berkeley, CA, 1980.
  119. Greenland, S., and Neutra, R. Control of confounding in the assessment of medical technology. *Intl. J. Epidemiol.* 9: 361-367 (1980).
  120. Sever, L. E. Hormonal pregnancy tests and spina bifida. *Nature* 242: 410-411 (1973).
  121. Lamm, S. H. Spontaneous abortions and forest spraying—model to test whether the data support the proposed association.

- Am. J. Epidemiol. 112: 438-439 (1980).
122. Miller, J. F., Williamson, E., Glue, J., Gordon, Y. B., Grudzin-skas, J. G., and Sykes, A. Fetal loss after implantation—a pro-spective study. *Lancet* ii: 554-556 (1980).
  123. Lancranjan, I., Popescu, H. I., Granvanescu, O., Klepsch, I., and Serbanescu, M. Reproductive ability of workmen occupa-tionally exposed to lead. *Arch. Environ. Health* 30: 396-401 (1975).
  124. Whorton, D., Krauss, R. M., Marshall, S., and Milby, T. H. Infertility in male pesticide workers. *Lancet* ii: 1259-1261 (1977).
  125. Rosenman, K. D., Anderson, H. A., Selikoff, I. J., Wolff, M. S., and Holstein, E. Spermatogenesis in men exposed to poly-brominated biphenyl (PBB). *Fertility Sterility* 32(2): 209-213 (1979).
  126. Milby, T. H., and Whorton, D. Epidemiological assessment of occupationally related, chemically induced sperm count suppres-sion. *J. Occup. Med.* 22: 77-82 (1980).
  127. Meyer, C. R. Semen quality in workers exposed to carbon di-sulfide compared to a control group from the same plant. *J. Occup. Med.* 23: 435-439 (1981).
  128. Wong, O., Utidjian, H. M. D., and Karten, V. S. Retrospective evaluation of reproductive performance of workers exposed to ethylene dibromide (EDB). *J. Occup. Med.* 21: 98-102 (1979).
  129. Levine, R. J., Symons, M. J., Balogh, S. A., Arndt, D. M., Kaswandik, N. T., and Gentile, J. W. A method for monitoring the fertility of workers. I. Method and pilot-studies. *J. Occup. Med.* 22: 781-791 (1980).
  130. Levine, R. J., Symons, M. J., Balogh, S. A., Milby, T. H., and Whorton, M. D. A method for monitoring the fertility of work-ers. II. Validation of the method among workers exposed to dibromochloropropane. *J. Occup. Med.* 23: 183-188 (1981).
  131. Harber, P. Some questions concerning "A Method for Monitor-ing the Fertility of Workers." *J. Occup. Med.* 23: 324 (1981).
  132. Dobbins, J. G., Eifler, C. W., and Buffler, P. A. The use of parity survivorship analysis in the study of reproductive out-come. Paper presented at Symposium on Methodologic and An-alytical Issues in Monitoring Human Populations for Reproductive Risks Associated with Environmental Exposures. Society for Epidemiologic Research Annual Meeting, Iowa City, Iowa, June 14-16, 1978.
  133. World Health Organization Scientific Group, Principles for the Testing of Drugs for Teratogenicity. World Health Organization Technical Report No. 364, 1967, pp. 1-18.
  134. Organization for Economic Cooperation and Development, Guideline for Testing of Chemicals—Teratogenicity. January, 1981.
  135. U.S. Environmental Protection Agency, Health Effects Test Guidelines, Chapter II. Specific Organ/Tissue Toxicity—De-velopmental Toxicity. Office of Toxic Substances. National Tech-nical Information Service, U.S. Department of Commerce, PB84-233295, 1984.
  136. U.S. Environmental Protection Agency. Pesticides Assessment Guidelines. Subdivision F. Hazard evaluation: human and do-mestic animals. EPA-540/98-2-025. National Technical Infor-mation Service, U.S. Department of Commerce, 1982.
  137. National Academy of Sciences, Principles and Procedures for Evaluating the Toxicology of Household Substances. Reproduc-tion and Teratogenicity Tests. Washington, DC, 1977, pp. 99-110.
  138. Interagency Regulatory Liaison Group, Recommended Guide-lines for Teratogenicity Studies in the Rat, Mouse, Hamster, or Rabbit, 1981.
  139. Armitage, P. Statistical Methods in Medical Research, Blackwell Scientific Publications, Oxford, 1971, pp. 181-184.
  140. Becker, B. A. The statistics of teratology. *Teratology* 9: 261-262 (1974).
  141. Haseman, J. K., and Hogan, M. D. Selection of the experimental unit in teratology studies. *Teratology* 12: 165-172 (1975).
  142. Kalter, H. The choice of the number of sampling units in tera-tology. *Teratology* 9: 257-258 (1974).
  143. Palmer, A. K. Statistical analysis and choice of sampling units. *Teratology* 10: 301-302 (1974).
  144. Staples, R. E., and Haseman, J. K. Selection of appropriate experimental units in teratology. *Teratology* 9: 259-260 (1974).
  145. Weil, C. S. Selection of the valid number of sampling units and a consideration of their combination in toxicological studies in-volving reproduction, teratogenesis or carcinogenesis. *Food Cosmet. Toxicol.* 8: 177-182 (1970).
  146. Weil, C. S. Choice of the number of sampling units in teratology. *Teratology* 10: 301 (1974).
  147. Searle, S. R. Linear Models. John Wiley and Sons, New York, 1971, pp. 376-420.
  148. Haseman, J. K., and Kupper, L. L. Analysis of dichotomous response data from certain toxicological experiments. *Biomet-rics* 35: 281-294 (1979).
  149. Williams, D. A. The analysis of binary responses from toxico-logical experiments involving reproduction and teratogenicity. *Biometrics* 31: 949-952 (1975).
  150. Kupper, L. L., and Haseman, J. K. The use of a correlated binomial model for the analysis of certain toxicological experi-ments. *Biometrics*, 34: 69-76 (1978).
  151. Gaylor, D. W. Methods and concepts of biometrics applied to teratology. In: *Handbook of Teratology, Volume 4: Research Procedures and Data Analysis* (J. G. Wilson and F. C. Fraser, Eds.), Plenum Press, New York, 1978, pp. 429-444.
  152. Gladen, B. The use of the jackknife to estimate proportions from toxicological data in the presence of litter effects. *J. Am. Statist. Assoc.* 74: 278-283 (1979).
  153. Lin, F. O., and Haseman, J. K. A modified Jonckheere test against ordered alternatives when ties are present at a single extreme value. *Biometr. Z.* 18: 623-631 (1976).
  154. Fears, T. R., Tarone, R. E., and Chu, K. C. False-positive and false-negative rates for carcinogenicity screens. *Cancer Res.* 37: 1941-1945 (1977).
  155. Gart, J. J., Chu, K. C., and Tarone, R. E. Statistical issues in interpretation of chronic bioassay tests for carcinogenicity. *J. Natl. Cancer Inst.* 62: 957-974 (1979).
  156. Gaylor, D. W. The use of safety factors for controlling risk. *J. Toxicol. Environ. Health* 11: 329-336 (1983).
  157. Krewski, D., and Van Ryzin, J. Dose response model for quantal response toxicity data. In: *Current Topics in Statistics and Prob-ability*, (Forgo, Dawson, Rao, and Selah, Eds.), North Holland Corp., Amsterdam, 1981.
  158. Van Ryzin, J. Estimating the mean of a random binomial pa-rameter with trial size random. *Sankhya* (Ser. B) 37: 10-27 (1975).
  159. Gaylor, D. W., and Kodell, R. L. Linear interpolation algorithm for low dose risk assessment of toxic substances. *J. Environ. Pathol. Toxicol.* 4: 305-312 (1980).
  160. Krewski, D., and Brown, C. Carcinogenic risk assessment: a guide to the literature. *Biometrics* 37: 353-366 (1979).